Komplexität einer künstlichen Intelligenz

vorgelegt von
Diplom-Informatiker Dr.-Ing.
Achim Hoffmann
aus Backnang

Vom Fachbereich Kommunikations- und Geschichtswissenschaft
der Technischen Universität Berlin
genehmigte Dissertation zur Erlangung des akademischen Grades

Doktor der Philosophie

Promotionsausschuß:
Vorsitzender: Prof. Dr. Karl Heinz Stahl
Berichter: Prof. Dr. Christoph Hubig
Berichter: Prof. Dr. Hans Poser
Berichter: Prof. Dr. Bernd Mahr (FB 20)

Tag der wissenschaftlichen Aussprache: 12. Juli 1993

Berlin 1993

D 83

A b s t r a c t

Hoffmann, Achim:

Komplexität einer künstlichen Intelligenz

Die Arbeit diskutiert die Möglichkeiten einer künstlichen Intelligenz (KI) vor dem Hintergrund ihrer Beschreibungskomplexität. Es werden philosophische Positionen, die mit Problemen der heutigen Forschung in der künstlichen Intelligenz in enger Beziehung stehen, unter dem eingeführten Komplexitätsblickwinkel analysiert.
Die Arbeit weist zunächst den allgemeinen Algorithmenbegriff als inadäquat für die Diskussion um die Möglichkeiten und Grenzen einer künstlichen Intelligenz zurück.
Stattdessen wird eine Einschränkung des allgemeinen Algorithmenbegriffs durch ein Beschreibungskomplexitätsmaß - den Begriff der algorithmischen Information oder - synonym - der Kolmogoroffkomplexität vorgeschlagen.
Darauf aufbauend wird zunächst der Zusammenhang zwischen aktuellen philosophischen Kritiken an der künstlichen Intelligenz (genauer an der *physical symbol system hypothesis*) und ihrer Beschreibungskomplexität erörtert. Hierbei geht es insbesondere um die auf Heideggers Phänomenologie gegründete Kritik an der KI von Dreyfus und Winograd & Flores. Einerseits wird sie als unzulänglich zurückgewiesen, andererseits wird die Kritik neu interpretiert, wobei sie spezifische Probleme der KI aufzeigt: Heideggers Phänomenologie weist auf menschliches Wissen, welches sich nicht durch Symbole repräsentieren läßt, die auf konkrete Gegenständlichkeiten verweisen.
Diese Erkenntnis wird mittels der Sprachphilosophie des späten Wittgenstein tiefergehend analysiert.
Ein Trugschluß wird bei der häufig durch philosophische Überlegungen (z.B. Wittgensteins Regelbegriff) begründeten Forderung nach subsymbolischen bzw. konnektionistischen Systemen aufgezeigt. Die Forderung ist nur *formal*, nicht aber in ihrer inhaltlichen Füllung gerechtfertigt. Formal gesehen, ist komplexes sinnvolles Verhalten von *zufälligem* Verhalten nicht zu unterscheiden. (Beides ist von hoher Kolmogoroffkomplexität.) Der Unterschied zeigt sich erst in der konkreten Ausprägung.
Weiterhin werden methodologische Probleme aufgezeigt, Prinzipien von Intelligenz zu finden, die wesentlich über die universelle Turingmaschine hinausgehen. Hier wird deutlich gemacht, daß eine Begründung des Allgemeinbegriffs von *Intelligenz* bzw. *Kognition* von entscheidender Bedeutung ist. Insbesondere wird hierbei auf die irreführende Rolle platonischer Ideen bei dem Entwurf von Experimenten und bei der Interpretation von anfänglichen Erfolgen einer neuen Technik hingewiesen.
Aus der eingeführten Komplexitätsperspektive wird die vieldiskutierte Frage nach den *prinzipiellen* Grenzen der künstlichen Intelligenz als unfruchtbar zurückgewiesen und durch eine fruchtbarere Frage ersetzt.

Inhaltsverzeichnis

Kapitel 1

Einleitung

Wie kann ein Etwas - ein Mensch, ein Tier oder ein Roboter - Erkenntnis über seine Umwelt durch seine Sinne erlangen und diese Erkenntnis dazu nutzen, um in irgendeiner Hinsicht zielorientiert zu agieren ?
Dies sind alte Fragen der Philosophie - sie könnten aber auch das Gegenstandsgebiet der künstlichen Intelligenz charakterisieren.
Insofern könnte man sagen, künstliche Intelligenz behandelt im großen Maßstab betrachtet philosophische Fragen.
Wichtige Fragen der täglichen Arbeit in der künstlichen Intelligenz sind beispielsweise: Was ist Bedeutung ? Was ist Rationalität ? Was ist vernünftiges Schließen ? Wie lassen sich Entscheidungen treffen und begründen ? etc.
Umgekehrt liegt es ebenfalls auf der Hand, daß philosophische Theorien zu den genannten Fragestellungen auch Eingang in die künstliche Intelligenz finden.

In der Philosophie ist die Frage nach der Natur von Bewußtsein, von Absichten, Intentionen, etc. und wie sie im Zusammenhang mit materiellen Gehirnsubstanzen oder gar mit künstlich intelligenten Wesen gedacht werden kann, von erheblichem Interesse.
Der Funktionalismus geht beispielsweise von der Annahme aus, daß jedem Bewußtseinszustand, jedem intentionalen Zustand ein entsprechender *funktionaler* Zustand des menschlichen Gehirns, bzw. eines anderen Trägersystems, korrespondiert.[1]
D.h. das 'Trägersystem' kann eine Menge von funktionalen Zuständen einnehmen, und zwar derart, daß die Zustandsübergänge innerhalb des 'Trägersystems' genau zu den Veränderungen der Bewußtseinszustände korrespondieren.
Somit will der Funktionalismus also Bewußtseinszustände durch die Angabe jeweils entsprechender Dispositionen, aus einem gegebenen Bewußtseinszustand in einen anderen Bewußtseinszustand überzugehen, bzw. bestimmtes Verhalten zu zeigen, erklären. Damit könnten Bewußtseinszustände auch durch ein entsprechendes Computerprogramm und einem je korrespondierenden Abarbeitungszustand erklärt werden. Fodor [Fod75] vertritt diesen Funktionalismus und zwar in der Form, daß er eine *angeborene* Lingua mentis

[1] Die Grundidee des sogenannten Funktionalismus geht auf Putnam zurück. Zu finden in mehreren Artikeln in Putnams Sammelwerk [Put75].

voraussetzt, die bereits alle je erforderlichen Elemente für eine mentale Repräsentation beinhaltet. Hierin geht Fodor weiter als Chomsky, der die Idee einer angeborenen universalen Tiefengrammatik für die menschliche Sprachfähigkeit vertritt.[2] Während Chomsky sich auf die rein syntaktischen Strukturen beschränkt, bezieht sich Fodor auf semantische Repräsentationen.
Putnam [Put91] führt unter anderem als Argumente gegen den Funktionalismus den 'Bedeutungsholismus'[3] an. Vor dem Hintergrund einer holistischen Sichtweise sei es schwer einzusehen, wie elementare semantische Einheiten für sich genommen eine spezifische Bedeutung haben sollen. Block [Blo78] führt eine Reihe von Argumenten an, die es unplausibel erscheinen lassen, daß sich Bewußtsein auf bloße funktionale Beziehungen gründen könnte und damit unabhängig vom menschlichen Gehirn möglich wäre. Hierbei wird insbesondere auf *Qualia* eingegangen, das sind Bewußtseinsphänomene wie *Schmerzempfindungen*, die mehr sein müssen, als nur die physischen Reaktionen, z.B. ein verzerrtes Gesicht. In dieser Arbeit soll die Bewußtseinsproblematik allerdings ganz ausgeklammert werden, da sie von keinerlei Bedeutung für die folgenden Betrachtungen ist.

Der Gegenstand der Arbeit ist die Komplexität der Beschreibung einer künstlichen Intelligenz[4] bzw. einer *deskriptiven* Theorie menschlicher Kognitionen. Für beide Fragestellungen spielt das Bewußtsein keine Rolle, da es lediglich um die *Beschreibung* und deren Eigenschaften von Intelligenzphänomenen geht, nicht aber um das Intelligenzphänomen selbst.

Der Wunsch, intelligente Wesen künstlich zu erzeugen, ist sehr alt; er läßt sich bis in die griechische Mythologie zurückverfolgen [McC79]. Aber erst durch die Erfindung des Computers, der die Konstruktionsmöglichkeiten von mechanischen Apparaturen in ihrer Vielfältigkeit unglaublich weit übertrifft, wurden die Voraussetzungen für eine künstliche Intelligenz geschaffen, für deren Durchführung bereits erste Pläne weit vorher (z.B. bei Babbage [Bab70]) existierten. Der Forschungsbereich der künstlichen Intelligenz - kurz KI - hat sich vor etwas mehr als drei Jahrzehnten konstituiert. Vielleicht ist hier Alan M. Turing mit seinem Artikel *Computing machinery and Intelligence* 1950 [Tur50] als einer der Pioniere zu nennen. Als eigentliche Geburtsstunde der heutigen künstlichen Intelligenz oder vielleicht genauer der *Artificial Intelligence* gilt die Dartmouth-Konferenz, die im Jahre 1956 stattfand. Seitdem haben zahlreiche Fachtagungen stattgefunden.

Auf den meisten internationalen Konferenzen zur künstlichen Intelligenz werden regelmäßig auch philosophische Beiträge vorgetragen, wodurch Einsichten in die Probleme der künstlichen

[2] Siehe hierzu Chomsky [Cho77].

[3] Vergleiche Abschnitt 8.2.

[4] Obgleich ein fehlendes Bewußtsein bei Maschinen häufig als Einschränkung oder Nachteil interpretiert wird, ist dies doch geradezu die moralische Rechtfertigung und damit auch der Antrieb für die wirtschaftliche Nutzung künstlich intelligenter Systeme. Die künstlich intelligenten Systeme sollen Arbeiten ausführen, die sonst von Menschen durchgeführt werden müßten. Andere Kulturen setzten Sklaven für derartige Arbeiten ein. Wenn Maschinen ein Bewußtsein zukommen würde - insbesondere wenn sie *Qualia* hätten, so würde man das moralisch begründete Ziel - den Einsatz von Maschinen statt Sklaven - nicht erreichen !

Intelligenz gewonnen werden sollen. Aber auch für die Philosophie konnten neue Einsichten aus der künstlichen Intelligenz gewonnen werden: Beispielsweise weist Frixione [Fri91] auf die Entwicklungen in der Logik hin, die in der KI von der antipsychologistischen Annahme der umfassenden deduktiven Rationalität hin zu Modellen führen, die Inkonsistenzen in großen Aussagensystemen zulassen, ohne daß dadurch jede Aussage als wahr abgeleitet werden kann, wie es bei der klassischen Logik der Fall ist. McCarthy [McC88] sieht die Möglichkeit, daß man in der KI so etwas wie eine Metaerkenntnistheorie entwickelt, die ähnlich wie die Metamathematik in der Mathematik, Aussagen über Möglichkeiten und Grenzen von bestimmten Erkenntnistheorien macht.[5] Aber auch in kleinerem Maßstab kann die KI philosophische Einsichten fördern. Beispielsweise weist Münch [Mün90] darauf hin, daß möglicherweise die Frühwerke Husserls erst durch die Entwicklungen der heutigen KI angemessen rezipiert werden können.

Die künstliche Intelligenz, der man große Erfolge schon für die 60er Jahre vorhersagte, stieß in den vergangenen Jahrzehnten jedoch auf größere Probleme, so daß ihre Erfolge weit unter den Erwartungen blieben. Allerdings gab es schon von Anbeginn Zweifel, ob menschliche Intelligenz überhaupt algorithmisch beschreibbar ist.
Schon vor der eigentlichen 'Geburtsstunde' der künstlichen Intelligenz, wurde die Frage nach ihren Grenzen diskutiert. Alan M. Turing, der 1937 die nach ihm benannte *Turingmaschine* entwickelte, erörterte 1950 in seinem bereits erwähnten Artikel *Computing machinery and intelligence* [Tur50] die Frage, ob Maschinen denken können. Die Turingmaschine wird allgemein als das theoretische Modell heutiger Computer betrachtet. Turing präzisierte in dem Begriff seiner Turingmaschine die intuitive Vorstellung dessen, was als durch Verfahrensregeln effektiv konstruierbar bzw. berechenbar gilt. Turing bemühte sich festzustellen, welche Prozesse sich durch eindeutige Regeln beschreiben lassen und damit *deterministisch* ablaufen. Ob die Aktivitäten des menschlichen Geistes, die menschliche Willensbildung oder das menschliche Denken deterministisch ablaufen, beschäftigte die Menschen allerdings schon sehr viel länger. Schon bei den Vorsokratikern, den frühen Atomisten, kam eine Diskussion um die Frage auf, ob der Mensch determiniert sei. In der heutigen Debatte um die Grenzen der künstlichen Intelligenz läßt sich zwar nicht genau diese Frage wiederfinden, es bestehen aber trotzdem sehr beachtliche Parallelen. Die alten Griechen hatten in erster Linie das Problem zu klären, wie eine deterministische Welttheorie, beispielsweise der Atomismus des Demokrit, mit dem wohl jedem Menschen vertrauten Gefühl der Freiheit sich für bestimmte Handlungen entscheiden zu können, vereinbaren läßt. Demgegenüber ist in der Debatte um die künstliche Intelligenz die Frage zu klären, ob wir das, was wir als Intelligenz bezeichnen, bei Maschinen wiederfinden können. Waren bis vor einigen Jahrzehnten Maschinen, die komplizierten Regeln deterministisch folgen noch nicht technisch realisierbar, so lassen sich heute und noch mehr in naher und fernerer Zukunft Maschinen konstruieren, die Millionen oder sogar Milliarden von Regeln folgen können. Nichtsdestotrotz gelang es dem Logiker Kurt Gödel bereits 1931 [Göd31] - also vor mehr als 50 Jahren - zu zeigen, daß es kein endliches formales

[5] In Chr. Hubig [Hub78] findet sich eine Analyse der Dialektik einer solchen Art.

System gibt, mit dessen Hilfe sich genau alle wahren Formeln der elementaren Arithmetik ableiten lassen. Diese Aussage zusammengenommen mit der allgemeinen Anerkennung der Turingmaschine als Modell beliebiger technisch realisierter Maschinen hat allerdings zur Folge, daß es auch nie eine Maschine geben wird, die genau alle wahren Sätze der elementaren Arithmetik konstruieren kann.
Ein Unterschied in der Fragestellung der griechischen Antike gegenüber der heutigen Diskussion um die künstliche Intelligenz liegt darin, daß man bei der KI nur von einer *endlichen* Zahl von Regeln, nach denen eine Maschine arbeitet, ausgeht. Die Griechen hingegen befaßten sich mit mit der metaphysischen Determinismusproblem, welches unabhängig von der Natur eines eventuell zugrunde liegenden Regelschemas zu behandeln ist. Daß die Tatsache, daß Maschinen nur endlich vielen Regeln folgen, nicht völlig unerheblich ist, zeigt genau das *Gödel*sche Ergebnis von 1931. Es gibt eben kein endliches Axiomensystem, und damit auch kein endliches Regelwerk und keinen Algorithmus, aus dem genau alle wahren Sätze der Arithmetik abgeleitet werden können. Aufgrund dieses Ergebnisses wurden in der Tat eine Reihe von Versuchen unternommen zu beweisen, daß eine künstliche Intelligenz - eine Maschinenintelligenz - von vornherein unmöglich ist.[6]

Auf der anderen Seite wurden in den letzten Jahren immer häufiger Argumente gegen die Möglichkeit einer künstlichen Intelligenz vorgebracht, die sich auf die Phänomenologie Heideggers gründen.
Heidegger setzte seine Phänomenologie der Alltäglichkeit der Phänomenologie seines Lehrers Edmund Husserl entgegen, die sich primär mit Erscheinungen in der wissenschaftlich-philosophischen Reflexion beschäftigte. Heidegger untersuchte hingegen, wie uns die Welt im alltäglichen Umgang begegnet. Damit kontrastierte er in *Sein und Zeit* die modelltheoretisch orientierte Sichtweise der 20er Jahre, die vielleicht am deutlichsten in Wittgensteins *Tractatus-logico-philosophicus* ihren philosophischen Ausdruck fand. Heidegger hingegen entwickelte seinen Begriff von *Zuhandenheit*. Die Zuhandenheit meint, daß Gegenständlichkeiten im alltäglichen Lebensvollzug gar nicht erst bewußt werden - gar nicht erst thematisiert werden - eben nur *zuhanden* sind. Im Gegensatz dazu steht bei Heidegger die *Vorhandenheit*. Die bewußte Reflexion einer konkreten Ontologie, von *vorhandenen* Gegenständen, wie es die modelltheoretische Vorstellung vorgibt, geschieht nur in besonderen Situationen.
Der weitaus unter den Prophezeiungen der sechziger Jahre zurückgebliebene Fortschritt der Forschung in der künstlichen Intelligenz und deren praktische Anwendungserfolge trugen erheblich dazu bei, daß immer mehr Stimmen laut wurden, die die generellen Möglichkeiten einer maschinellen Intelligenz in Zweifel zogen. An dieser Stelle kam die Philosophie Heideggers als tiefsinnige Fundierung der Kritik sehr willkommen. Sie wurde von Dreyfus bereits 1965 [Dre65] in diesen Zusammenhang gebracht - zu einer Zeit, in der sie zunächst wenig beachtet wurde.
Im wesentlichen wird Heideggers *Sein und Zeit* in diesem Zusammenhang so interpretiert, daß Heidegger nachwies, daß das menschliche Denken nicht stereotypen Regeln folgt; ja

[6] Siehe z.B. Lucas [Luc61, Luc70]. Siehe hierzu auch Abschnitt 9.5.

daß das menschliche Denken sich nicht einmal an festen ontologischen Vorstellungen orientiert. Beides wird hingegen Maschinen als inhärente Eigenschaft zugeschrieben, - sie müssen stereotypen Regeln folgen und dabei mit einer festen Repräsentation einer Ontologie arbeiten und unterscheiden sich bereits allein darin fundamental vom menschlichen Denken.
Im folgenden werden dreierlei Regelbegriffe verwendet werden: Der *algorithmische* Regelbegriff wird in Kapitel 3 formal eingeführt. Im Gegensatz dazu wird der *philosophische* Regelbegriff verwendet werden, welcher in Abschnitt 8.1 erläutert wird. Der Terminus der *stereotypen Regel* soll als vorwissenschaftlicher Begriff des algorithmischen Regelbegriffs verstanden werden.

Auch neuere erkenntnistheoretische Sichtweisen, wie Wittgensteins Spätphilosophie, die sich in seinen *Philosophischen Untersuchungen* niederschlug, oder damit eng verwandte Sichtweisen, wie der Quine'sche Holismusgedanke, betonen die enorme Komplexität menschlicher Begriffs- bzw. Sprachverwendung. Dadurch, durch die genannten phänomenologisch motivierten Kritiken an der traditionellen künstlichen Intelligenz und durch einige technische Neuerungen, kam in den 80er Jahren ein Trend in der künstlichen Intelligenz auf, der bereits Ende der 50er Jahre schon einmal eine Blüte erlebte.
Neuronale Netzwerke, heute oft auch als *konnektionistische Systeme* bezeichnet, geben aufgrund der genannten philosophischen Sichtweisen Hoffnung zu einem neuen, erfolgreicheren Weg zu einer künstlichen Intelligenz.
Diese Forschungsrichtung, der Konnektionismus, scheint vielen der in der philosophischen Erörterung deutlich gewordenen Eigenschaften menschlichen Denkens zu entsprechen. Hier wird eine künstliche Intelligenz angestrebt, die sich gänzlich von dem symbolischen, dem programmorientierten Ansatz zu einer künstlichen Intelligenz entfernt.
Die Unzulänglichkeiten der symbolischen künstlichen Intelligenz werden im konnektionistischen Ansatz vielfach als überwunden angesehen. Smolensky [Smo88] und andere sehen daher den Konnektionismus auch als adäquate theoretische Grundlage einer Kognitionswissenschaft.
Die Kognitionswissenschaft, als die Wissenschaft vom menschlichen Denken, muß natürlich die philosophischen Erkenntnisse noch eher berücksichtigen, als eine künstliche Intelligenz, die im wesentlichen daran interessiert ist, das Ergebnis menschlichen Denkens nachzubilden. Schließlich gelingt es Flugzeugen auch auf andere Weise zu fliegen, als ihren natürlichen Vorbildern - den Vögeln. Insofern könnte man sich eine künstliche Intelligenz, die auf ganz andere Weise zu intelligenten Leistungen kommt als der Mensch, noch vorstellen, doch eine Kognitionswissenschaft soll das spezifisch menschliche Denken beschreiben.
Wie bereits erwähnt, hatte die künstliche Intelligenz weit geringere Erfolge zu verzeichnen, als man in den 60er Jahren vorhersagte. Dabei hatte man in den darauffolgenden Jahrzehnten zahlreiche Ansätze zur Modellierung menschlicher Intelligenzleistungen entwickelt. Viele Ansätze erschienen anfänglich durchaus vielversprechend. Doch ließ sich generell die Tendenz feststellen, daß sich die Ansätze weitaus weniger nutzbringend für die

Lösung sogenannter 'real-world problems'[7] einsetzen liessen, als für die ursprünglich konzipierten 'Testprobleme'[8] mit denen die Ansätze während ihrer Entwicklungszeit geprüft wurden.

In der kontrovers diskutierten Frage nach den Grenzen der künstlichen Intelligenz nahm die Arbeit ihren gedanklichen Ursprung.[9] Der Frage also, ob menschliche Denkprozesse algorithmisch sind ? Beide Positionen erscheinen plausibel:

- Die unbegrenzten Möglichkeiten einer künstlichen Intelligenz erscheinen plausibel, da zu jeder faktisch aufgezählten Menge - und damit endlichen Menge - von geforderten Intelligenzleistungen ein geeigneter Algorithmus angebbar ist.
- Andererseits scheint aufgrund folgender Überlegung die Begrenztheit einer maschinellen Intelligenz plausibel: In jedem faktisch angegebenen Algorithmus läßt sich eine Intelligenzleistung - zumindest im Prinzip angeben - die der Algorithmus nicht hervorzubringen vermag.[10]

Beide Positionen gehen von unterschiedlichen Voraussetzungen aus: Im ersten Fall werden die hervorzubringenden Intelligenzleistungen fixiert und ein geeigneter Algorithmus angegeben. Im zweiten Fall wird ein Algorithmus fixiert und eine Intelligenzleistung gefordert, die der Algorithmus nicht erfüllen kann.
Diese Problematik läßt sich - solange man endliche Mengen von Intelligenzleistungen betrachtet, nicht lösen. Dies liegt an der *Inadäquatheit* des Algorithmenbegriffs für diese Fragestellung, wie er in seiner formalen Definition für mathematische Zwecke und potentielle Unendlichkeiten in der bekannten Form[11] vorliegt. Im Bereich von Intelligenzleistungen, die auch von Menschen hervorgebracht werden können, - insbesondere wenn man an bestimmte Anwendungen wie medizinische Diagnose, Konstruktionsaufgaben etc. denkt, bewegt man sich zunächst immer im endlichen Bereich. Eine unendliche Menge von intelligenten Leistungen eines Menschen muß Spekulation bleiben, solange Menschen nur eine endliche Zeit leben. Für eine Diskussion um die Grenzen der KI auf der Basis von endlichen Mengen von Intelligenzleistungen muß auch der Algorithmenbegriff entsprechend eingeschränkt werden. Genauer - die Länge eines zu betrachtenden Algorithmus darf nicht beliebig sein, sondern muß durch eine Maximalzahl von Regeln, oder eine ma-

[7] Als 'real-world problems' werden in der künstlichen Intelligenz Computeranwendungen von praktischer Relevanz bezeichnet. Diese stehen im Gegensatz zu den häufig sehr einfachen 'Testproblemen', an denen man einen neuen Ansatz im Labor prüft. Es hat sich herausgestellt, daß viele Ansätze im Labor bei den ihnen gestellten Aufgaben erfolgreich sind, während die Ansätze im Einsatz für praktisch relevante Aufgaben versagen.

[8] Die 'Testprobleme' sind Beispielprobleme für neue Ansätze in der künstlichen Intelligenz, an denen geprüft werden soll, ob ein Ansatz erfolgversprechend ist. Solch ein 'Testproblem' könnte beispielsweise das erfolgreiche Spielen des Mühlespiels sein. Vergleiche auch die vorhergehende Fußnote.

[9] Auch hier war die Bewußtseinsfrage ausgeklammert.

[10] Beispielsweise nach dem Schema des Gödelschen Unvollständigkeitstheorems. Vergleiche hierzu Abschnitt 9.5.

[11] Siehe Kapitel 3 für Einzelheiten.

ximale Programmlänge eingeschränkt sein. Dann erst kann man sinnvoll über Grenzen einer künstlichen Intelligenz sprechen.
Somit entwickelte sich die Arbeit aus der Neubetrachtung vieler Fragestellungen zum Thema einer möglichen künstlichen Intelligenz bzw. der algorithmischen Natur menschlicher Denkprozesse auf der Basis des modifizierten Algorithmenbegriffs. Dies bezieht sich nicht nur auf philosophische Kritiken an der KI, wie der von Dreyfus, sondern geht über metaphorische Nachbildungen des menschlichen Gehirns durch künstliche neuronale Netze oder der Selbstorganisation eines kognitiven Systems bis zum Begriff der Kreativität.
Eine besondere methodologische Problematik in der KI und der Kognitionswissenschaft, die aufgrund der nun in den Vordergrund getretenen Endlichkeit von Mengen von zu fordernden Intelligenzleistungen bzw. von zu beschreibenden Denkprozessen deutlich wurde, wird im zweiten Teil der Arbeit behandelt.
Zu der genannten Einschränkung des zu betrachtenden Algorithmenbegriffs bot sich der formale Begriff der Kolmogoroffkomplexität an. Der Begriff der Kolmogoroffkomplexität befaßt sich mit der Beschreibungskomplexität von Zeichenketten, welche zur Beschreibung von Intelligenzphänomenen genutzt werden können. Hierbei konnte gezeigt werden, daß sich in einem gewissen *absoluten* Sinn von der Komplexität einer Zeichenkette sprechen läßt. Die Kolmogoroffkomplexität mißt dabei die *kürzestmögliche* Beschreibung einer Zeichenkette. Ursprünglich wurde dieser Begriff von A. N. Kolmogoroff entwickelt, um einer mathematischen Fassung von *Zufall* näher zu kommen.[12]
Die Kolmogoroffkomplexität kann auch als Maß für die Zahl von Regeln aufgefaßt werden, die notwendig sind, um gegebene Phänomene - z. B. konkrete Intelligenzleistungen - zu beschreiben oder zu erklären.
Aus der neuen Perspektive der Kolmogoroffkomplexität wird in der Arbeit zunächst eine methodologische Erklärung für das häufige Scheitern der vielen Ansätze zu einer künstlichen Intelligenz an 'real-world problems' gegeben. Für die Diskussion um die Adäquatheit des symbolischen bzw. des konnektionistischen Ansatzes für die künstliche Intelligenz bzw. für die Kognitionswissenschaft ermöglicht die Perspektive neue grundlegende Einsichten. Erkenntnistheoretische und sprachphilosophische Positionen werden unter dem Komplexitätsaspekt in Beziehung zu einer möglichen künstlichen Intelligenz gesetzt. Letztlich wird noch die Frage nach den Grenzen der künstlichen Intelligenz mittels des Begriffs der Kolmogoroffkomplexität einer scharfen Präzisierung zugeführt.

Die Arbeit ist wie folgt aufgebaut:

Im ersten Teil wird in die künstliche Intelligenz, in die Grundbegriffe der theoretischen Informatik sowie in Ideen zu einer Grundlegung der Kognitionswissenschaft eingeführt. Dazu wird im folgenden Kapitel zunächst ein grober Überblick über die künstliche Intelligenz gegeben. Im dritten Kapitel wird auf die theoretischen Grundlagen von deterministisch arbeitenden Maschinen eingegangen werden. Am Ende dieses Kapitels wird der zentrale Begriff dieser Arbeit, der Begriff der algorithmischen Information bzw. der Kol-

[12]Eine *zufällige* unendliche Folge von Ereignissen, z.B. die Augenzahl beim Würfeln, ist dadurch gekennzeichnet, daß sie von unendlicher Kolmogoroffkomplexität ist.

mogoroffkomplexität eingeführt. In Kapitel 4 werden verschiedene Positionen zur Frage der Repräsentation von Wissen dargestellt sowie ein Versuch der Fundierung der Kognitionswissenschaft vorgestellt.

Im zweiten Teil der Arbeit werden wissenschaftstheoretische Betrachtungen zu einer Methodologie der künstlichen Intelligenz und Kognitionswissenschaft angestellt. Zu diesem Zweck wird zunächst im fünften Kapitel die philosophische Diskussion um das Universalienproblem kurz referiert. Die Frage nach dem Ursprung von Universalien erscheint für eine angemessene Eingrenzung des zu untersuchenden Phänomenbereichs in der künstlichen Intelligenz und Kognitionswissenschaft erforderlich. Im darauffolgenden Kapitel wird auf methodologische Probleme einer Wissenschaft eingegangen, die versucht komplexe Phänomene durch generelle Prinzipien zu beschreiben. Hier wird deutlich werden, daß eine scharfe Abgrenzung des Allgemeinbegriffs von *Intelligenz* bzw. von *Kognitionen* von besonderer Bedeutung ist. In der Tat läßt sich ein methodologischer Zirkel aufweisen, der bei der gängigen Herangehensweise in zumindest einem Teilgebiet der künstlichen Intelligenz besteht:

> *Durchgeführte Fallstudien in Experimentalwelten bestätigen im wesentlichen lediglich, was bei der Konzeption der Fallstudie bereits implizit angenommen wurde.*

Dies wird am Beispiel des maschinellen Lernens ausgeführt.

Im dritten Teil der Arbeit wird auf die Zusammenhänge des in der Arbeit eingenommenen Komplexitätsblickwinkels mit verschiedenen neueren philosophischen Positionen eingegangen.
Dabei wird zunächst der Zusammenhang zwischen aktuellen philosophischen Kritiken an der (symbolischen) künstlichen Intelligenz und dem Begriff der Kolmogoroffkomplexität erörtert. Hierbei geht es insbesondere um die phänomenologische Kritik an der künstlichen Intelligenz von Dreyfus und Winograd & Flores.
Kapitel acht behandelt sprachphilosophisch-erkenntnistheoretische Positionen in diesem Zusammenhang. Ein Trugschluß wird bei der häufig durch philosophische Überlegungen (z.B. Wittgensteins Regelbegriff) begründeten Forderung nach subsymbolischen bzw. konnektionistischen Systemen aufgezeigt. Die Forderung ist nur *formal*, d.h. äußerlich, nicht aber in ihrer inhaltlichen Füllung gerechtfertigt. Formal gesehen, ist komplexes sinnvolles Verhalten von *zufälligem* Verhalten nicht zu unterscheiden. (In beiden Fällen ist die Verhaltensbeschreibung von hoher Kolmogoroffkomplexität.) Der Unterschied zeigt sich erst im konkreten Inhalt; *welches* Verhalten in *welcher* Situation gezeigt wird.
Im neunten Kapitel wird die vieldiskutierte Frage nach den *prinzipiellen* Grenzen der künstlichen Intelligenz erörtert und durch eine Einschränkung präzisiert.
Hierbei werden auch die Grenzen der in der Arbeit vorgelegten Argumentation - nämlich die Begrenzung auf das Turingmaschinenmodell - diskutiert. Dem Argument, daß mögliche physikalische Systeme mit analogen und/oder asynchronen Signalen nicht durch Turing-

maschinen zu modellieren sind, und mithin eine Argumentation basierend auf dem Turingmaschinenmodell hinfällig ist, wird ein komplexitätstheoretisches Argument entgegengesetzt. Argumente gegen die Möglichkeit einer KI die sich auf das menschliche Bewußtsein berufen, werden ebenfalls behandelt.
Im Schlußkapitel werden die in der Arbeit neu aufgeworfenen Gesichtspunkte noch einmal kurz angesprochen. Es werden Konsequenzen der in der Arbeit vorgestellten Überlegungen für die Philosophie sowie für die Forschung in der künstlichen Intelligenz aufgezeigt. Mögliche weitergehende Forschungsarbeiten werden umrissen sowie deren potentielle Konsequenzen für die Philosophie skizziert.

Teil I

Grundlagen

Der erste Teil dieser Arbeit dient einer allgemeinen Einführung in Themengebiete, die aus einer neuen Perspektive betrachtet werden sollen. Dabei wird verhältnismäßig viel Platz für die Darstellung von Gebieten verwendet, die nicht so eng mit der traditionellen Philosophie zusammenhängen.
Zunächst wird in Kapitel 2 ein allgemeiner Überblick über die verschiedenen Arbeitsfelder der heutigen künstlichen Intelligenz gegeben. Dieser Überblick erhebt keinen Anspruch auf Vollständigkeit - er soll lediglich einen Eindruck vermitteln, für welchen Typ von Fragestellungen die künstliche Intelligenz technische Lösungen anzubieten sucht.

In Kapitel 3 wird der in den 30er Jahren präzisierte Algorithmenbegriff behandelt. Dabei wird sowohl die gedankliche Entwicklung des Begriffs dargestellt, als auch seine formale Fassung durch den Begriff der Turingmaschine. Ausgehend vom Turingmaschinenbegriff werden *universelle* Turingmaschinen erklärt. Auf universellen Turingmaschinen letztlich basiert der für die Arbeit so zentrale Begriff der *Kolmogoroffkomplexität* oder - synonym zu verwenden - der Begriff der *algorithmischen Information*.

Kapitel 4 widmet sich häufig diskutierten Aspekten der künstlichen Intelligenz. Dabei geht es zunächst in Abschnitt 4.1 um den möglichen Zusammenhang zwischen Symbolmanipulation, menschlichem Denken und Intentionen. In diesem Zusammenhang wird auch Pylyshyns Konzeption einer Grundlegung der Kognitionswissenschaft vorgestellt, welche menschliches Denken als Symbolmanipulation im *buchstäblichen* Sinne ausweisen soll. Die kontrovers diskutierte Frage, ob menschliches Wissen durch Symbole adäquat repräsentiert werden kann, wird in Abschnitt 4.2 zusammen mit einer Darstellung der wichtigsten Positionen vorgestellt.

Kapitel 2

Die künstliche Intelligenz

Sind wir intelligent genug, um Intelligenz zu verstehen ? Ein Ansatz diese Frage zu beantworten, ist die 'künstliche Intelligenz' (kurz auch KI), ein Teilgebiet der Informatik, das sich damit auseinandersetzt, wie intelligente Maschinen zu konstruieren sind. Unter anderem versucht man auf dem Gebiet der künstlichen Intelligenz Maschinen zu konstruieren, die Probleme lösen, Spiele spielen, bestimmte Muster erkennen und klassifizieren, mathematische Sätze beweisen, natürliche Sprache verstehen und sogar eine gewisse Lernfähigkeit zeigen, indem sie ihr Verhalten dahingehend ändern, daß sie gleiche oder ähnliche Aufgaben bei einer Wiederholung in irgendeiner Hinsicht besser oder schneller lösen.
Der Forschungsbereich der künstlichen Intelligenz hat sich vor etwas mehr als drei Jahrzehnten konstituiert. Alan M. Turing schrieb 1950 einen Artikel mit dem Titel *Computing machinery and Intelligence* [Tur50] und könnte somit als einer der ersten KI Wissenschaftler genannt werden. Als eigentliche Geburtsstunde der heutigen künstlichen Intelligenz oder - vielleicht genauer - der *Artificial Intelligence* gilt jedoch die Dartmouth-Konferenz, die im Jahre 1956 stattfand. Zu den Teilnehmern der damaligen Konferenz zählten unter anderem John McCarthy, Marvin Minsky, Allen Newell[1] und Herbert Simon, die auch heute noch zu den führenden Köpfen der künstlichen Intelligenz gehören.
Die enge Beziehung zwischen künstlicher Intelligenz und Informatik liegt in dem gemeinsamen Interesse, Methoden zur effektiven Verwendung des Computers zu entwickeln und zu benutzen. So kann man die künstliche Intelligenz als Teilgebiet der Informatik betrachten, aber auch umgekehrt die Informatik als Teilgebiet der künstlichen Intelligenz. Die KI hat neben ihrer engen Beziehung zur Informatik auch noch Verbindungen mit der Philosophie, der Linguistik, der Kognitionswissenschaft und der Psychologie.
Die Ziele der KI lassen sich im wesentlichen in zwei Punkte unterteilen (Winston [Win92]):

a) Die Konstruktion von Maschinen, die ein im allgemeinen als intelligent bezeichnetes Verhalten hervorbringen (engineering approach).

b) Durch Simulation von kognitiven Prozessen, die menschliche Intelligenz erforschen (cognitive approach).

[1] Allen Newell ist 1992 verstorben.

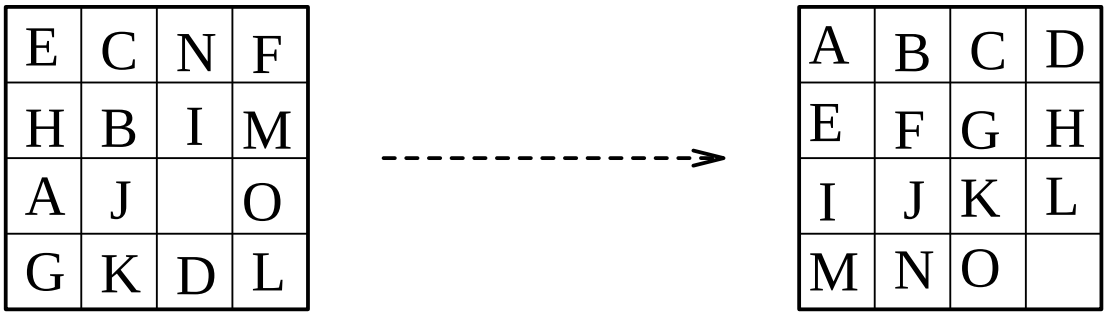


Abbildung 2.1: Das 15-Puzzle. Ein Standardbeispiel für ein schwieriges Suchproblem der künstlichen Intelligenz.

Im folgenden sollen einzelne Forschungsgebiete der KI kurz erläutert werden:

2.1 Suchen und Problemlösen

Unter diesem Punkt werden in der KI Suchprobleme subsumiert. Das sind Probleme, deren Lösung aus einer meist sehr großen Zahl von potentiellen Lösungen bestimmt werden muß. Als Beispiel sei das 15-Puzzle genannt, bei dem in einem 4x4 Felderquadrat 15 durchnumerierte oder durchbuchstabierte verschiebbare Plättchen angeordnet sind. Siehe hierzu Abbildung 2.1. Ein Plättchen kann nur in das (einzige) leere Feld des 4x4 Quadrats entweder vertikal oder horizontal verschoben werden. Aus einer beliebigen Anfangsanordnung der Plättchen ist durch geeignetes Verschieben der Plättchen in das jeweils leere Feld eine vorgegebene Plättchenanordnung zu erzielen (z.B. in alphabetischer Reihenfolge geordnet). Bei einem solchen Problem gibt es allein mehr als 10 000 000 000 000 verschiedene Anordnungen der Plättchen. Daraus resultieren noch wesentlich mehr verschiedene Möglichkeiten, Plättchen aus einer gegebenen Anfangssituation heraus zu verschieben. Die Aufgabe nun, für eine Anfangssituation eine Verschiebestrategie zu bestimmen, die durch möglichst wenig Verschiebungen die vorgegebene Zielkonstellation erzeugt, erfordert - von einem Menschen durchgeführt - sicherlich Intelligenz.
Aber auch für heutige Computer ist die Zahl der verschiedenen Schiebestrategien viel zu groß, um sie alle nacheinander zu prüfen und festzustellen, ob es sich jeweils um eine mögliche Lösung handelt. Solche als *kombinatorische Probleme* bezeichneten Aufgaben tauchen in vielerlei Gestalt auf. Beispielsweise tritt ein solches kombinatorisches Problem auch bei dem automatischen Beweisen von mathematischen Sätzen auf. Hier ist die Frage zu lösen, welche logischen Ableitungsregeln in welcher Reihenfolge auf welche Formeln angewendet werden müssen, um überhaupt eine - oder besser noch eine einigermaßen kurze - Ableitungskette zu erhalten.
Obwohl ein Computer sicherlich wesentlich schneller verschiedene Ableitungsmöglichkeiten bzw. Verschiebungsmöglichkeiten bei dem 15-Puzzle untersuchen kann, ist die menschliche Intelligenz dem Computer auf sehr vielen Gebieten - zumindest bis heute noch - weit überlegen. Der Grund hierfür ist darin zu sehen, daß die menschliche Intelligenz nicht

primär die verschiedenen Möglichkeiten 'blind' durchspielt, sondern vielmehr von vornherein eher 'strategisch geeignete' Entscheidungen trifft, in welcher Richtung die Suche nach einer Lösung unter den vielen potentiellen Lösungen fortgesetzt werden soll. Dieses selektive Suchen nach einer Lösung, wodurch bei weitem nicht mehr alle Möglichkeiten, sondern nur noch ein kleiner Bruchteil durchgespielt werden muß, versucht man auch auf Computer zu übertragen. Nichtsdestotrotz ist dies, wie oben erwähnt, bislang nicht in ausreichendem Maße gelungen. Dies ist ein Punkt, an dem Kritiker der KI wie Dreyfus einhaken und argumentieren, daß Computer nie in der Lage sein werden, eine letztlich so effektive Lösungssuche durchzuführen, wie es von Menschen getan wird.

2.2 Spiele spielen

Dieses Gebiet der KI befaßt sich damit, Programme zu entwickeln, die Spiele erfolgreich spielen, bei denen man allgemein sagt, daß sie Intelligenz erfordern würden. Das populärste Spiel dieser Art ist sicherlich Schach. Obwohl in der Aufgabe eine gute Schachpartie zu spielen, kein besonderer Nutzen zu sehen ist, wird dieses Problem in der KI aus den folgenden Gründen doch ernst genommen:
Das Schachspiel ist hinreichend kompliziert, so daß der Mensch eine Vielzahl verschiedener Aspekte seiner Intelligenz einsetzt. Bei Spielen wie Schach ist die Aufgabe mit all ihren Randbedingungen einschließlich dem Ziel klar definiert. Dadurch kann man sich, so erhofft man, besser auf die eigentlichen Prinzipien, die den menschlichen Problemlösungsprozeß zugrunde liegen, konzentrieren.
Die Schachprogrammierung läßt sich als *Drosophila-Fliege der künstlichen Intelligenz* bezeichnen - es ist nur ein Versuchsfeld auf dem Erkenntnisse gesammelt und Techniken entwickelt werden sollen, die sich dann (hoffentlich) auf andere wichtige Gebiete der KI übertragen lassen.
Seit dem Alan M. Turing 1953 [Tur53] das erste Schachprogramm schrieb, das er durch eine Handsimulation[2] demonstrierte, sind in der Schachprogrammierung große Fortschritte erzielt worden. Beim Schachspiel ebenso wie bei vielen anderen Einsatzgebieten von 'intelligenten' Programmen, wurde in der Anfangsphase der KI die vorhandene kombinatorische Explosion, das heißt, die Vielzahl der bei der Lösungssuche zu betrachtenden Möglichkeiten, bei weitem unterschätzt. In den Abschnitten 8.6 und 9.1 wird noch etwas näher auf den heutigen Stand der Schachprogrammierung eingegangen.

2.3 Automatisches Beweisen

In diesem Zweig der künstlichen Intelligenz wurden in erster Linie Methoden entwickelt, um die bereits erwähnte kombinatorische Explosion möglichst gering zu halten. Methoden mit Hilfe derer man nach mechanischen, also programmierbaren Verfahren logisch

[2]Eine Handsimulation eines Programms ist die gedankliche Ausführung des Programms mit Bleistift und Papier.

korrektes Schließen nachvollziehen kann, wurden bereits vor der Entwicklung der ersten elektronischen Rechenmaschine von Mathematikern wie Hilbert, Gentzen, Skolem und anderen entwickelt.[3] Diese Verfahren stellten sich jedoch für den praktischen Gebrauch in Computerprogrammen als zu wenig effizient, das heißt, zu zeitraubend heraus. Als eine der bekanntesten Entwicklungen auf dem Gebiet des automatischen Beweisens ist das Resolutionsverfahren von Robinson [Rob65] zu nennen. Dies ist ein Beweisverfahren mit nur einer einzigen Ableitungsregel oder besser einem einzigen Ableitungsregelschema. Allerdings handelt es sich dabei weniger um eine Nachbildung des menschlichen logischen Schließens, als um eine speziell für Maschinen entwickelte Methode.

2.4 Verstehen natürlicher Sprache

Auf diesem Gebiet versucht man Programme zu entwickeln, die gesprochene oder geschriebene Sprache, wie sie als Verständigungsmittel zwischen Menschen gebraucht wird, zu 'verstehen'. Der Terminus 'verstehen' hat schon vielfach in der nicht nur philosophischen Literatur Grund zum Anstoß gegeben. Hier wird häufig behauptet, daß Maschinen grundsätzlich nicht dazu in der Lage sind, etwas zu verstehen. Daß Computer mit formalisierten Sprachen wie der Sprache der Prädikatenlogik oder einer Programmiersprache umgehen können, steht schon lange außer Zweifel. Das 'Verstehen' natürlicher Sprache jedoch hat bisher erheblich mehr Schwierigkeiten in sich geborgen als man zu Anfang erwartet hatte. Wurde Ende der fünfziger Jahre doch noch euphorisch damit gerechnet, daß man in wenigen Jahren Maschinen hätte, die in der Lage sind, Texte einer Sprache in eine andere Sprache - z.B. Englisch in Deutsch - zu übersetzen. Als die Versuche nicht den erwarteten Erfolg zeitigten, wurde man bescheidener. Es stellte sich heraus, daß für die Übersetzung eines Textes in eine andere Sprache die Analyse der grammatikalischen Struktur der einzelnen Sätze in Verbindung mit einem Wörterbuch bei weitem nicht ausreicht. Beispielsweise müssen Zusammenhänge, die sich über den gesamten Text erstrecken können, richtig eingeordnet werden. Referenzen können in bestimmten Fällen nicht aufgrund der grammatikalischen Satzstruktur, sondern nur aufgrund der Inhalte richtig aufgelöst werden. Als die amerikanischen Regierung Bar-Hillel damit beauftragte, ein Gutachten über die Erfolgsaussichten solcher Anstrengungen zu schreiben, wurden in den USA, als Folge von Bar-Hillels Gutachten [BH64], die Gelder für Forschungsaufträge auf diesem Gebiet radikal gestrichen.
Nichtsdestotrotz erwachten die Versuche Systeme zu bauen, die natürliche Sprache verarbeiten, zu neuem Leben, als in den 70er Jahren Terry Winograd mit seiner Dissertation über Sprachverarbeitung einen neuen Ansatz ausarbeitete [Win72]. Zur Demonstration seines Ansatzes in seinem System SHRDLU wählte er eine Mikrowelt, in der es eine beschränkte Anzahl von verschiedenfarbigen, unterschiedlich großen und unterschiedlich geformten Holzblöcken gab, die in einem abgegrenzten Raum beliebig angeordnet werden konnten. Das System sollte einen Roboterarm steuern, der die einzelnen Blöcke bewegen

[3] In der Tat arbeitete bereits Leibniz an einem solchen Kalkül, mit dem er das logische Schließen durch Rechnen ersetzen wollte.

konnte. Nun konnten Fragen an das System gestellt werden, die sich auf geometrische Lagebeziehungen zwischen einzelnen Blöcken bezogen. Dem System konnten auch Anweisungen in natürlicher Sprache gegeben werden, wie es die Position der Blöcke verändern solle. Bemerkenswert war dabei, daß das System nicht nur die Anweisungen, die in einem Satz gegeben wurden, ausführte, sondern es stellte auch Mehrdeutigkeiten in Anweisungen fest und versuchte solche Mehrdeutigkeiten durch das Herstellen von Bezügen zu dem bisherigen 'Gesprächsverlauf' aufzulösen. Die Richtigkeit dieser Schlüsse ließ sich das System vom Benutzer jeweils bestätigen. Winograds Arbeit ließ selbst Bar-Hillel an seinem früheren Gutachten zweifeln.
In der Folge stellte sich jedoch heraus, daß sich bei der Erweiterung von Winograds Blockwelt auf umfangreichere Anwendungsbereiche erhebliche und bisher unüberwindliche Schwierigkeiten einstellten, welche hauptsächlich auf das sehr umfangreiche, zusätzlich notwendige 'Wissen' zurückzuführen sind. Mittlerweile hat Winograd selbst Abstand von seinem Ansatz genommen und vertritt die Ansicht, daß der Versuch, Systeme zu entwickeln, die natürliche Sprache verarbeiten, unfruchtbar sei [WF86].

2.5 Bildverstehen

Dieses Gebiet hat sich heute eigentlich als eigenständige Forschungsdisziplin innerhalb der Informatik etabliert. Beispielsweise sollen auf diesem Gebiet Systeme entwickelt werden, die den Fertigungsablauf in einer Fabrik überwachen können. Dazu ist es notwendig, die einzelnen Werkstücke auf einem Fließband sicher voneinander unterscheiden zu können. Es ist zwar heute nicht mehr schwierig mit entsprechenden Kameras hochaufgelöste farbige Bilder zu gewinnen, die dann akurat wiedergegeben werden können. Versuche, ein solches Bild zu interpretieren, Gegenstände, die aus unterschiedlichster Perspektive und unter unterschiedlichsten Beleuchtungsbedingungen dargestellt sein können, zu identifizieren, stossen jedoch auf ganz erhebliche Schwierigkeiten. Noch schwieriger ist es, bewegte Bilder, das heißt Bildfolgen von einer kleinen Szenerie adäquat zu interpretieren. Wenn sich in der Szenerie Gegenstände bewegen, und sich dadurch veränderte Lichtspiegelungen und Schatten ergeben, so impliziert dies noch weitere Probleme.

2.6 Maschinelles Lernen

Vielleicht ist das maschinelle Lernen Träger der größten Hoffnungen unter den Teilgebieten der KI. Es hat einen ähnlichen Werdegang wie die Sprachverarbeitung. Als die ersten Computer verfügbar waren, lag auch die Vorstellung in der Luft, lernende Systeme zu bauen, die sich anschließend selbsttätig an ihre Umwelt anpassen und sich alles Notwendige zur Erfüllung ihrer Aufgaben durch selbständiges Lernen aneignen. Es sollte also ein Elementarsystem entwickelt werden, das im wesentlichen nur allgemeine Lernstrategien beinhaltet. Dies sollten Regeln sein, nach denen die Eingaben des Systems - beispielsweise die Meldung, ob die gerade bearbeitete Aufgabe erfolgreich durchgeführt wurde - geeignet weiterverarbeitet werden, so daß die Problemlösungsfähigkeit des Systems sich

zunehmend erweitert und verbessert. Zu den ersten Lernsystemen, die mit solchen Zielen entwickelt wurden, zählt Samuels Checker Player [Sam59, Sam67], ein Programm für das Damespiel, das sich durch gespielte Partien selbst verbessert. In der Tat gelang es einer später verfeinerten Version dieses Systems, auf menschlichem Meisterniveau zu spielen. Ein anderes frühes Lernsystem, das einfache Lernaufgaben bewältigte, ist Rosenblatts Perceptron [Ros59]. Ein System, das in seiner Struktur an Nervennetze erinnert, wobei sich bestimmte interne Werte in den einzelnen Netzknoten verändern können. Beide Ansätze liessen sich in der Folge jedoch nicht auf andere, wichtigere Gebiete übertragen. Man kam auch auf dem Gebiet der lernenden Systeme - ähnlich wie in der Sprachverarbeitung - zu der Überzeugung, daß sich auch durch weitere Forschungsaktivitäten kein Erfolg einstellen würde.
Knapp zehn Jahre später, 1970 trat Patrick Winston mit seiner Dissertation über ein lernendes System an die Öffentlichkeit [Win70]. Winstons System war in der Lage eine Klassifikationsregel für einen aus Bauklötzen zusammengestellten Torbogen zu lernen. Das System lernte die Klassifikationsregel, indem ihm positive und negative Beispiele für einen solchen Torbogen vorgelegt wurden, d.h. es wurden Konstellationen vorgelegt, die einem Torbogen entsprachen und andere, die keinem Torbogen entsprachen. Welche der vorgelegten Beispiele Torbögen waren, wurde dem System mit den Beispielen mitgeteilt. Winstons Arbeit erweckte die versiegten Forschungsanstrengungen auf dem Gebiet der lernenden Systeme zu neuem Leben. Seit dem wurden mehr und mehr Arbeiten zu diesem Thema veröffentlicht. Seit Mitte der 80er Jahre wurden aufgrund neuer Entwicklungen auch Ansätze, die auf der Nachbildung von Neuronennetzen basieren, wieder populär.

2.7 Expertensysteme

Durch die Entwicklung von Expertensystemen - Systemen, die auf einem eingeschränkten Gebiet einen Experten vertreten oder zumindest bei der Problemlösung unterstützen sollen - wurde das Interesse an lernenden Systemen noch verstärkt. Solche Expertensysteme haben beispielsweise die Aufgabe, aufgrund bestimmter Krankheitssymptome eine Krankheit zu diagnostizieren. Es gelang tatsächlich für eine eingeschränkte Zahl von Krankheiten geeignete Systeme zu entwickeln. Hier sei das bekannte System MYCIN genannt, das für die Klassifikation bestimmter Blutkrankheiten eingesetzt werden kann [BS84]. Bei der Entwicklung solcher Expertensysteme stellte sich heraus, daß es häufig sehr schwierig ist, an das Wissen heranzukommen, das ein Experte auf seinem Gebiet hat. Das heißt, in diesem Fall die Regeln, nach denen er seine Diagnose bestimmt. Der gebräuchliche Weg besteht darin, daß ein sogenannter Wissensingenieur mit verschiedenen Methoden, beispielsweise durch geeignete Führung eines Interviews,[4] versucht, das Wissen des Experten in eine programmierbare Form zu bringen.
An dieser Stelle zeigen sich erhebliche Schwierigkeiten. Einerseits fällt es einem Experten oft schwer, Regeln für seine Entscheidungen anzugeben [DD87]. Dieser Schwierigkeit versucht man unter anderem durch Forschungsanstrengungen in der Kognitionspsychologie

[4] Die eingesetzten Methoden stammen zum Teil aus der kognitiven Psychologie.

Herr zu werden [MS88]. Andererseits gibt es dabei auch viele praktische Schwierigkeiten zu überwinden; darunter fällt die Tatsache, daß Experten häufig wenig Zeit für die detaillierte Erklärung ihrer Problemlösungsmethoden haben. Oft fehlt es auch an geeigneten Wissensingenieuren. Insgesamt wird durch diese Situation die Entwicklung eines Expertensystems zu einer kostspieligen Angelegenheit, da der Wissenstransfer, also die Interaktion des Experten mit dem Wissensingenieur und die Umsetzung der gewonnenen Erkenntnisse in ein Programm durch den Wissensingenieur, ziemlich zeitaufwendig ist.
Aus diesem Grund ist die Idee von lernenden Systemen besonders interessant für die Konstruktion von Expertensystemen. Vielleicht ist es hierbei möglich, ein System zu entwickeln, das den Experten selbständig befragt, wobei es möglicherweise sogar noch gezieltere Fragen stellen könnte, als ein menschlicher Wissensingenieur und anschließend das vom Computer ausführbare Programm automatisch erstellt.[5] Ein lernendes System könnte auch ganz darauf verzichten, einen Experten zu befragen und sich darauf beschränken, einen Experten bei der Problemlösung zu beobachten: Zum Beispiel zu beobachten, welche Diagnose der Experte bei welchen Symptomen stellt und daraus durch induktive Schlüsse zu den richtigen Diagnoseregeln zu gelangen.

2.8 Konnektionismus und neuronale Netzwerke

In den letzten Jahren hat der Konnektionismus, der sich mit neuronalen Netzwerken befaßt, einen grandiosen Zuwachs an Popularität erhalten. Unter anderem wurde dieser Aufschwung durch die Arbeiten von Rumelhart et al. [RMt86] stimuliert. Konnektionistische Systeme basieren auf der Idee, eine Vielzahl von Verarbeitungseinheiten zu einem System zuammenzufassen. Diese vielen Verarbeitungseinheiten sollen untereinander stark vernetzt sein. Dadurch beeinflussen sich die einzelnen Verarbeitungseinheiten ständig gegenseitig in ihrer Funktion. Siehe hierzu Abbildung 2.2. Insofern ist der Konnektionismus weniger als ein Teil der künstlichen Intelligenz in dem Sinn zu sehen, daß es einen abgegrenzten Aufgabenbereich dafür gibt. Vielmehr stellt er einen Ansatz dar, mit dem ein großer Teil der KI-Teilgebiete erfolgreich behandelt werden soll; dazu zählen u.a. die Sprachverarbeitung, die Bildverarbeitung und das Lernen.
Die Idee erfreut sich unter anderem wegen der augenscheinlichen Verwandtschaft zur neurobiologischen Organisation des menschlichen Gehirns, daß offensichtlich zu intelligentem Verhalten in der Lage ist, großer Beliebtheit. Erste Arbeiten zu künstlichen neuronalen Netzen wurden bereits in den 40er Jahren von McCulloch & Pitts [MP43] und von Hebb [Heb49] durchgeführt. Auch heute sind die damals von Hebb vorgeschlagenen elementaren Lernregeln noch aktuell und werden bei der Entwicklung neuronaler Systeme eingesetzt. Besonders durch Rosenblatts Perceptron [Ros59] wurden neuronale Systeme gegen Ende der fünfziger Jahre populär. Dieser ersten Hoffnungswelle wurde durch Minskys und Paperts Buch *Perceptron* [MP69] ein fast völliger Abbruch beschert. Minsky und Papert untersuchten mit rein mathematischen Methoden die Möglichkeiten und Grenzen

[5] Einen guten Überblick über die verschiedenen Ansätze zum maschinellen Lernen findet man in Michalski et al. [Mic90] oder in Shavlik & Dietterich [SD90].

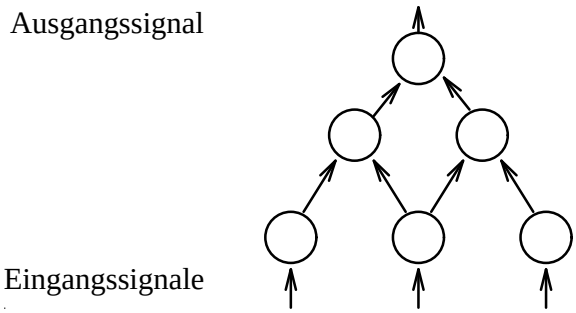


Abbildung 2.2: Schema eines kleinen konnektionistischen Systems. Es handelt sich um ein sogenanntes 3-Lagennetz, da drei Verarbeitungseinheiten auf dem Pfad vom Eingangs- zum Ausgangssignal hintereinander geschaltet sind.

von neuronalen Systemen. Sie kamen dabei zu dem Ergebnis, daß derartige neuronale Systeme bereits an sehr einfachen Aufgaben aus bestimmten strukturinhärenten Gründen scheitern müssen. Sie beschränkten ihre Analyse allerdings auf sogenannte Zweilagennetze. Dies sind Netze, bei denen höchstens zwei Verarbeitungseinheiten hintereinander geschaltet sind. Der Grund, den sie für die Einschränkung ihrer Untersuchungen anführten, bezog sich auf das folgende Problem: Bei *Mehrlagennetzen* sei es nicht zu sehen, daß lokale Lernregeln[6] - die zu untersuchen waren, überhaupt jemals zu einem adäquaten Lernergebnis führen können.

Die Proponenten des Konnektionismus führen in erster Linie die folgenden Gründe an:

- Die Verarbeitung verläuft hochgradig parallel. Dies könnte einen Beitrag zur Bewältigung der bei KI-Aufgaben typischerweise vorhandenen kombinatorischen Komplexität leisten.
- Die Systemstruktur ist an menschliche oder biologische Informationsverarbeitungsmodelle angelehnt, welche bereits erwiesen haben, daß sie eines intelligenten Verhaltens fähig sind.
- Die herkömmlichen Ansätze zur KI, die sogenannten symbolischen Ansätze, haben nicht die erwarteten Erfolge gezeigt.
- Die konnektionistischen Systeme arbeiten nicht mehr auf einer symbolischen Ebene, sondern auf einer *subsymbolischen* Ebene. Das heißt, man kann keiner einzelnen Verarbeitungseinheit oder einer kleinen Gruppe solcher Einheiten eine bestimmte Bedeutung zuordnen, wie man es bei bestimmten Einträgen in einem Wissensrepräsentationssystem könnte. Hingegen ergibt sich bei konnektionistischen Systemen das Verhalten erst durch ein Zusammenwirken aller oder zumindest sehr vieler

[6] Lokale Lernregeln sind Lernregeln, die den internen Zustand einer Verarbeitungseinheit nur aufgrund der lokal zu beobachtenden Signale bestimmen, dies sind die Ein- und Ausgangssignale der jeweiligen Verarbeitungseinheit und gegebenenfalls noch deren Nachbareinheiten, sowie des momentanen internen Zustands der Verarbeitungseinheit.

Verarbeitungseinheiten des Systems. Der innere Zustand einer einzelnen Verarbeitungseinheit wird durch die inneren Zustände vieler anderer Verarbeitungseinheiten beeinflußt und beeinflußt seinerseits viele andere Einheiten. Dies scheint der in den letzten Jahren immer mehr beachteten Kritik von Dreyfus [DD87, DD88] und Winograd & Flores [WF86] aus einer phänomenologischen Perspektive entgegen zu kommen.

- Man verspricht sich eine sehr gute Lernfähigkeit dieser Systeme durch die Möglichkeit, schrittweise und verteilt in einer Vielzahl von Verarbeitungseinheiten Begriffe zu lernen. Weiterhin wird behauptet, daß konnektionistische Systeme besonders wenig empfindlich gegenüber leicht fehlerhaften Lerndaten seien. Dies scheint wiederum dem menschlichen Begriffslernen näher zu kommen, als die traditionellen symbolischen Ansätze zum maschinellen Lernen.

Mittlerweile gibt es eine kaum noch zu übersehende Vielfalt von Vorschlägen zum konkreten Aufbau von konnektionistischen Systemen. Es werden oft zu den Vorschlägen spezifische Probleme genannt, die im Experiment mit dem vorgeschlagenen konnektionistischen System gelöst werden konnten.
Kritisch zu bemerken ist hierbei allerdings, daß die jeweils gelösten Aufgaben häufig von einer erstaunlichen Einfachheit sind.

Nun wurden die wichtigsten Teilgebiete der künstlichen Intelligenz kurz angerissen. Insgesamt läßt sich sagen, daß die mehr pragmatisch orientierten Forscher der KI, die überwiegende Mehrheit der Forschergemeinschaft bilden. Dies sind diejenigen, die zunächst einmal nur nützliche Systeme entwickeln wollen, statt schwierig zu modellierende Phänomene menschlicher Intelligenz nachzubilden.

Kapitel 3

Effektive Berechenbarkeit und die formalen Grundlagen von Maschinen

Der Begriff der *effektiven Berechenbarkeit*,[1] konnte erst in den 30er Jahren unseres Jahrhunderts präzisiert werden. Das damalige Interesse an diesem Begriff war weniger durch eine Entwicklung heutiger Computer motiviert, sondern vielmehr durch erkenntnistheoretische Fragen zur Grundlegung der Mathematik. Man suchte nach einer Präzisierung dessen, was sich tatsächlich vom menschlichen Geist *konstruieren* läßt. Durch den Unendlichkeitsbegriff[2] in der Mathematik wurde diese Frage wichtig: Welche Elemente von unendlichen Mengen lassen sich konkret erzeugen - d.h. irgendwann, nach Anwendung von einer *endlichen* Zahl von Konstruktionsschritten ? Insofern expliziert der Begriff der effektiven Berechenbarkeit die Fähigkeit des menschlichen Geistes, konkrete Objekte zu konstruieren, welche hierbei durch syntaktische Ausdrücke beschrieben werden. Da jede endliche Menge sich immer in einer Liste notieren läßt, liegt das Interesse natürlicherweise erst bei der Konstruktion unendlicher Objektmengen.[3] Die Bedeutung des Begriffs der effektiven Berechenbarkeit, auch *Algorithmus* genannt, geht über den Bereich der reinen Mathematik hinaus. Der Begriff präzisiert nicht nur die Klasse von effektiven Berechnungsverfahren, sondern auch die Mächtigkeit von wissenschaftlichen Beschreibungsspra-

[1] Der Begriff der *effektiven Berechenbarkeit* wurde ab 1936 auf verschiedene Weisen unabhängig voneinander expliziert. Dazu zählt unter anderem die bereits erwähnte Arbeit von A. M. Turing [Tur37], der Begriff der allgemein rekursiven Funktionen von A. Church [Chu36b] und die sogenannten *Postschen Kalküle* von E. L. Post [Pos43]. Es ließ sich jedoch in der Folge nachweisen, daß all diese Explikationen bezüglich der mit ihnen berechenbaren Funktionen äquivalent sind. Diese Feststellung führte zur Erhärtung der sinngemäß von Church, Post und Turing unabhängig voneinander formulierten und nach erstem benannten Churchschen These, daß der intuitive Begriff der effektiven Berechenbarkeit tatsächlich durch den Begriff der allgemein rekursiven Funktionen bzw. der Turingmaschine expliziert wird. Seitdem ist es trotz einer ganzen Reihe von ernsthaften Versuchen nicht gelungen, einen darüber hinausgehenden Berechenbarkeitsbegriff zu explizieren. Daher gilt die Churchsche These heute als sehr gut bestätigte Annahme. Siehe hierzu auch Davis [Dav82].

[2] Genauer genommen war es nicht nur *ein* Unendlichkeitsbegriff; es ging um abzählbar unendliche und um überabzählbare Mengen, um potentielle und aktuale Unendlichkeiten.

[3] Für eine weitergehende Darstellung der damaligen Fragestellung siehe auch Körner [Kö68] oder Wang [Wan74].

chen, innerhalb derer durch formale Mittel Einzelaussagen abgeleitet werden sollen. Die Problematik, die dabei für die Forschungsmethodologie in der künstlichen Intelligenz und der Kognitionswissenschaft erwächst, wird in Kapitel 6 ausführlich behandelt.
Im folgenden Abschnitt wird die historische Entwicklung des Algorithmenbegriffs dargestellt. Im zweiten Abschnitt werden Aspekte des Algorithmenbegriffs hervorgehoben, die einer allgemeinen Betrachtung von Algorithmen dienlich sein sollen. Im dritten Abschnitt wird der Begriff der *universellen Turingmaschine* vorgestellt. Im letzten Abschnitt des Kapitels wird auf den vorhergehenden Abschnitten aufbauend der Begriff der algorithmischen Information oder - synonym - der Kolmogoroffkomplexität eingeführt, der für die vorliegende Arbeit von zentraler Bedeutung ist.

3.1 Die Entwicklung des Algorithmenbegriffs

Insbesondere in der Mathematik und Logik hat man schon seit langem versucht, die entwickelten Berechnungs- und Schlußverfahren möglichst präzise zu beschreiben[4]. Dort spricht man statt von Verfahren häufig auch von Algorithmen[5]. Der Begriff des Algorithmus wurde erst im 20. Jahrhundert soweit präzisiert, daß auch mechanische Apparaturen einen Algorithmus ausführen können. Ein Algorithmus ist eine endliche Menge von Regeln, die uns genau angeben, was in jedem Moment der Ausführung eines solchen Algorithmus zu tun ist, um eine bestimmte Klasse von Aufgaben oder Problemen zu lösen.[6] Wenn ein Algorithmus zur Lösung einer Klasse K von Problemen gegeben ist, so ist jedermann dazu in der Lage, jedes einzelne Problem dieser Klasse K zu lösen. Vorausgesetzt, er ist in der Lage die einzelnen Operationen, die der Algorithmus vorschreibt, auszuführen. Die einzelnen Operationen werden durch die Regeln, aus denen sich der Algorithmus zusammensetzt, bestimmt. So kann beispielsweise ein Schuljunge den euklidischen Algorithmus[7] erlernen und ihn richtig anwenden, ohne dabei zu verstehen, warum der Algorithmus zu einer richtigen Antwort führt.
Ein Mensch, der einen Algorithmus anwendet, benutzt dafür bestimmte Daten. Darunter fallen die Regeln des Algorithmus, die er im Kopf hat und die übrigen Daten, die er verwendet und die außerhalb seines Kopfes festgehalten sind - zum Beispiel auf Papier niedergeschrieben. Man kann also davon ausgehen, daß Speichermedien benutzt werden, um einerseits Daten, die vor der Ausführung eines Algorithmus vorliegen, zu speichern. Dazu zählen die Regeln des Algorithmus selbst und die Beschreibung des zu lösenden

[4] Siehe zum Beispiel Arbeiten von Leibniz [Bur80], Frege [Fre79] oder Hilbert & Bernays [HB68].

[5] Der Name 'Algorithmus' ist an den frühen arabischen Gelehrten *Ibn Mûsâ Al-Chwârismî* angelehnt, der bereits im 9. Jahrhundert ein Buch über die *Regeln der Wiedereinsetzung und Reduktion* schrieb.

[6] Beispielsweise könnte eine solche Klasse aus der Addition zweier beliebiger gegebener Zahlen bestehen. Einen dafür geeigneten Algorithmus hat jeder in der Schule gelernt.

[7] Der euklidische Algorithmus wurde von Euklid (etwa um 300 v.Chr.) als Verfahren zur Bestimmung des größten gemeinsamen Teilers (ggT) zweier natürlicher Zahlen a und b angegeben: Dabei wird 1. der Rest r der Division $a : b$ gebildet. Gibt es keinen Rest, so ist b der ggT. Ansonsten wird 2. a durch b und b durch r ersetzt. 3. Gehe zu Schritt 1. Daß der Algorithmus immer zu dem richtigen Ergebnis kommt, ist nicht unmittelbar einzusehen.

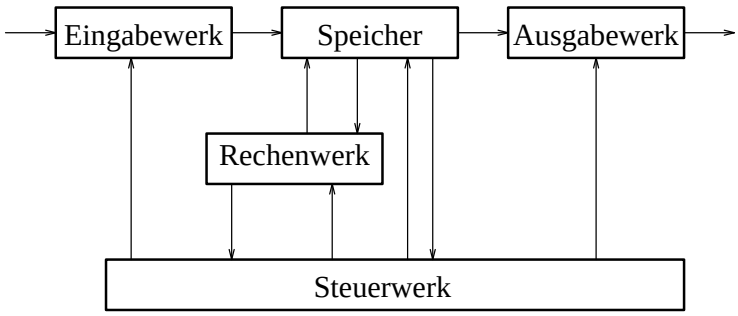


Abbildung 3.1: Das Schema der von Neumann Rechnerarchitektur.

Problems. Andererseits wird ein Speichermedium benötigt, um Zwischenergebnisse zu speichern, die im Laufe der Ausführung eines Algorithmus entstehen. Zweitens muß dieser Mensch eine festumrissene Menge von elementaren Operationen ausführen können. Nämlich genau diejenigen Operationen, die in den verschiedenen Regeln des Algorithmus vorkommen. Als drittes ist zu beachten, daß die Reihenfolge und die Auswahl der Regeln und der in ihnen beinhalteten Operationen richtig durchgeführt wird.
Von diesen drei Aufgaben ist die zweite Aufgabe, die richtige Ausführung von elementaren Operationen, sicherlich am einfachsten technisch zu realisieren - insbesondere, wenn die elementaren Operationen hinreichend einfach gehalten werden. Die erste Aufgabe, das Speichern von Daten - insbesondere des Algorithmus - erscheint schon schwieriger, da es auch nicht so offensichtlich ist, wie die *allgemeine Form* einer algorithmischen Regel aussieht. Die dritte Aufgabe, die Regeln des Algorithmus in der richtigen Reihenfolge auszuwählen, muß ebenfalls für *beliebige* Algorithmen gelöst werden. Sie muß also auch für beliebige Regelmengen richtig funktionieren.
Wenn man die praktische Realisierung von Rechenmaschinen betrachtet, so ist die dritte Aufgabe in Form eines sogenannten Steuerwerks realisiert. Die zweite Aufgabe übernimmt eine sogenannte arithmetisch-logische Einheit. Bei der ersten Aufgabe wurde bei den frühen Entwicklungen von Computern der Algorithmus außerhalb des Einflußbereichs der Maschine durch externe Knöpfe eingestellt. In späteren Entwicklungen von Rechenmaschinen wurde der Algorithmus zusammen mit den Zwischenergebnissen, die bei der Ausführung des Algorithmus entstehen, in einem elektrischen oder elektronischen Speicher abgelegt. Dies ist eine grobe Beschreibung des architektonischen Prinzips der meisten heutigen Computer, die von Neumann-Architektur (Abbildung 3.1).
Der Begriff des Algorithmus war bis in die dreißiger Jahre dieses Jahrhunderts eher vage. Was beispielsweise eine Regel ist, die mechanisch interpretiert werden kann und ohne zusätzliche Hintergrundannahmen korrekt von einer mehr oder weniger einfachen Maschine ausgeführt werden kann, ist zunächst unklar. Zu diesem Zweck wird in jedem Fall eine spezifische Sprache benötigt, die einerseits hinreichend ausdrucksstark sein muß, um in ihr die Regeln des Algorithmus ausdrücken zu können, und andererseits darf es bei deren Interpretation keine Zweifelsfälle und Mehrdeutigkeiten geben. So gelangt man von dem Problem, den Begriff eines Algorithmus zu explizieren, zu zwei ähnlich schwierigen Problemen: Es muß eine 'mechanische' Sprache zusammen mit einer Maschine definiert werden,

die in der Lage ist, die Sätze der mechanischen Sprache eindeutig und korrekt zu interpretieren. Die Maschine muß geeignete elementare Operationen ausführen. Zu diesem Zweck könnte man eine große Zahl von Algorithmen sammeln und sie auf ihre Gemeinsamkeiten hin untersuchen. Hierbei ist es allerdings nicht klar, wie eine solch induktive Vorgehensweise zu einer zufriedenstellenden Abstraktion von den konkret betrachteten Algorithmen führen soll.
Alan M. Turing [Tur37] betrachtete Algorithmen, die von Menschen korrekt durchgeführt werden können. Dadurch kam er zu einer Reihe von elementaren Operationen, die mit Sicherheit von einem mechanischen Apparat richtig ausgeführt werden können. Andererseits reichten sie dafür aus, auch die schwierigsten Algorithmen, die Menschen ausführen können, durch einen mechanischen Apparat ausführen zu lassen. Turings elementare Operationen können zu komplexeren Operationen zusammengesetzt werden. Dabei benötigt man nur entsprechend mehr Anweisungsregeln, also einen entsprechend längeren Algorithmus und mehr Speicherplatz für die entstehenden Zwischenergebnisse. Wenn eine solch elementare Maschine erst einmal konzipiert ist, dann ist es nicht mehr schwierig eine passende Sprache zu entwerfen, in der sich Regeln formulieren lassen, die von der Maschine eindeutig interpretiert werden können.
Wenn man sich einen Menschen vorstellt, der auf einem in Quadrate unterteiltem Stück Papier eine Berechnung durchführt, so spielen die folgenden Dinge dabei eine wesentliche Rolle:

1. Ein Speichermedium; das Stück Papier.

2. Eine Sprache, die Symbole beinhaltet, um Zahlen und Bewegungsrichtungen des Schreibstiftes darzustellen.

3. Einen abgegrenzten Bereich des Speichermediums, der gerade von dem berechnenden Menschen betrachtet wird, das heißt bestimmte Kästchen des Papiers.

4. Geisteszustände, d.h. zu jedem Zeitpunkt der Berechnung merkt sich der Mensch in welchem Stadium der Berechnung er sich befindet. Daraufhin entscheidet er, welches der nächste Berechnungsschritt ist.

5. Die Ausführung des nächsten Berechnungsschritts, der aus den folgenden Komponenten besteht:

   a) Die Veränderung der Symbole, die auf den betrachteten Kästchen stehen. Dies kann durch bloßes Einschreiben eines Symbols in ein bisher leeres Feld geschehen. Falls auf dem fraglichen Kästchen bereits ein Symbol steht, so wird dieses ausradiert und durch ein neues Symbol ersetzt.
   b) Das Augenmerk wird auf eine andere abgegrenzte Menge von Kästchen gelegt.
   c) Der Mensch nimmt einen anderen 'Geisteszustand' ein.

Was diesen Vorgang mechanisch macht, sind die beiden folgenden Eigenschaften:

1. Das Prinzip der Determiniertheit.
2. Das Prinzip der Endlichkeit.

Entsprechend dem ersten Prinzip hängt die Entscheidung, welcher Berechnungsschritt als nächster durchgeführt wird, nur von den Symbolen auf den abgegrenzten Kästchen auf dem Papier und dem gegenwärtigen 'Geisteszustand' ab. Wenn eine Berechnung gemäß einem vorgegebenen Algorithmus durchgeführt wird, so muß der Algorithmus für jede Situation, in die der Berechnende während der Berechnung gelangt, angeben, was zu tun ist. Eine solche Situation wird dabei durch die Symbole, die in den betrachteten Kästchen auf dem Papier stehen und durch den spezifischen 'Geisteszustand' in dem sich der Berechnende befindet, charakterisiert.

Der hier benutzte Begriff 'Geisteszustand' bedarf dabei einer genaueren Erklärung. Zunächst soll zur größeren Klarheit die Wahrnehmung des Berechnenden eines Algorithmus ohne Verlust der Allgemeinheit in zwei getrennte Teile unterteilt werden: Der eine Teil seien die visuellen Wahrnehmungen, das heißt, genau die Symbole, die der Berechnenden auf denjenigen Kästchen seines Papierstückes sieht, die seinem momentanem Augenmerk gelten. Der andere Teil seiner momentanen Wahrnehmungen bezieht sich auf seinen inneren Sinn. Dazu gehören sicherlich auch viele irrelevante Dinge. Beispielsweise könnte er feststellen, daß sein Magen knurrt; dies sollte jedoch keinerlei Einfluß darauf haben, welcher Berechnungsschritt als Nächstes durchgeführt wird. Vielmehr können wir davon ausgehen, daß die für die korrekte Ausführung eines Algorithmus relevanten Wahrnehmungen durch den inneren Sinn sich ausschließlich darauf beziehen, welche Berechnungsschritte bisher bereits durchgeführt wurden.
Damit kann das, was als 'Geisteszustand' bezeichnet wurde, wie folgt präzisiert werden: Es handelt sich um eine eindeutige Disposition, auf bestimmte Symbole, die auf dem Papier in den betrachteten Kästchen stehen, in genau festgelegter Weise zu reagieren; also den nächsten Berechnungsschritt auszuwählen. In dem 'Geisteszustand' soll also lediglich gespeichert werden, an welcher Stelle des gesamten Berechnungsverfahrens sich der Berechnende augenblicklich befindet.
Gemäß dem oben genannten Prinzip der Endlichkeit soll davon ausgegangen werden, daß der Berechnende nur in der Lage ist, eine endliche Anzahl von verschiedenen 'Geisteszuständen' in obigem Sinne einzunehmen. Gleichermaßen wird davon ausgegangen, daß der Berechnende lediglich in der Lage ist, zu jedem Zeitpunkt eine endliche Zahl von unterschiedlichen Zeichen in den Kästchen auf dem Papier wahrzunehmen. Ja, man kann sogar davon ausgehen, daß sowohl die Zahl der verschiedenen 'Geisteszustände', als auch die Zahl der unterschiedlichen Zeichen, die wahrgenommen werden können, eine feste Obergrenze haben. Die Obergrenze für die unterschiedlichen Symbole auf dem Papier wird verhältnismäßig klein sein, während die Obergrenze für die Zahl der verschiedenen 'Geisteszustände' wesentlich größer sein wird. Daraus folgt unmittelbar, daß der Berechnende zu jedem Zeitpunkt auch nur eine endliche Zahl von Kästchen auf dem Papier überblicken kann. Wenn es notwendig ist, mehr Kästchen in Betracht zu ziehen, so muß dies in mehreren Schritten zeitlich nacheinander geschehen.

Aus den obigen Erörterungen folgt, daß die Zahl der unterschiedlichen Zeichen, die in ein Kästchen geschrieben werden können, endlich ist. Dies folgt einerseits aus der endlichen Zahl von eindeutigen Dispositionen, den 'Geisteszuständen' und andererseits hat es bei der Ausführung eines Algorithmus wenig Sinn, Zeichen auf das Papier zu schreiben, die anschließend nicht mehr eindeutig klassifiziert werden können. Man kann sich zwar vorstellen, prinzipiell unendlich lange Zeichenketten zu schreiben, dies würde aber wiederum zeitlich nacheinander durch wiederholtes Schreiben einzelner Zeichen aus einer endlichen Auswahl von Zeichen geschehen. Nun soll das Prinzip der Endlichkeit noch auf die Zahl der Operationen angewendet werden, die in jedem Rechenschritt durchgeführt werden können. Wie oben bereits erwähnt, besteht ein solcher Rechenschritt aus dreierlei Typen von Operationen:

1. Der Berechnende geht in Abhängigkeit der Zeichen in den betrachteten Kästchen und seinem augenblicklichen 'Geisteszustand' entweder in einen anderen 'Geisteszustand' über, oder er bleibt in dem bisherigen Zustand.

2. Gegebenenfalls werden Zeichen in einigen oder in allen der betrachteten Kästchen ausradiert und durch andere ersetzt. Die Zahl der Zeichen ist aber in jedem Fall endlich. Der Einfachheit wegen kann man ohne Beschränkung der Allgemeinheit davon ausgehen, daß in jedem Berechnungsschritt nur genau ein Kästchen bearbeitet wird.[8]

3. Als dritter Operationstyp bleibt noch die Verlagerung des Augenmerks auf andere Kästchen des benutzten Papiers. Auch hier kann man davon ausgehen, daß das Augenmerk sich nur um eine endliche Zahl von Kästchen in einer von einer endlichen Zahl von Richtungen verlagert. Der Vereinfachung wegen kann man davon ausgehen, daß das Augenmerk in einem einzelnen Berechnungsschritt immer nur um ein Kästchen in eine der vier Richtungen verschoben werden kann. Als weitere Vereinfachung kann schließlich noch die Zweidimensionalität des Speichermediums, das heißt des Schreibpapiers, aufgegeben werden. Stattdessen kann man die Betrachtungen auf ein eindimensionales Band ohne Verlust der Allgemeinheit beschränken.[9]

[8] Daß dies keine Einschränkung bedeutet, kann man wie folgt sehen: Für eine endliche Zahl k von Kästchen, in denen jeweils eines von m verschiedenen Zeichen steht (das leere Feld ist auch als eines dieser Zeichen zu betrachten) gibt es höchstens m^k, also endlich viele verschiedene Möglichkeiten, diese Kästchen mit Zeichen zu beschreiben. Mithin können die k Kästchen auch als ein einziges Kästchen betrachtet werden, in das allerdings entsprechend mehr unterschiedliche Zeichen eingeschrieben werden können. Andererseits läßt sich auch durch folgende Betrachtung sehen, daß die Einschränkung auf ein einzelnes Feld keine Einschränkungen für einen auszuführenden Algorithmus bedeuten: Anstatt in einem Berechnungsschritt in mehrere Kästchen andere Zeichen einzuschreiben, kann dieser Berechnungsschritt auch in mehrere elementarere Berechnungsschritte zerlegt werden, die ihrerseits nacheinander durchgeführt werden.

[9] Dies ist möglich, da alle Kästchen in der Ebene auf die Kästchen auf einem eindimensionalen Band abgebildet werden können. Dazu muß das Band nur entsprechend lang sein und die Berechnungsregeln des Algorithmus müssen geeignet abgeändert werden. Es läßt sich sogar zeigen, daß jede Berechnung mit einem mehr als zweidimensionalen Speichermedium ebenso mit Hilfe eines solchen Bandes durchgeführt

Zusammenfassend kann man also davon ausgehen, daß der Berechnende zu jedem Zeitpunkt der Berechnung immer nur genau ein Kästchen auf dem eindimensionalen Band betrachtet. Als Berechnungsoperationen nur den Inhalt genau dieses Kästchens verändert und das Augenmerk im Anschluß daran entweder auf dem gleichen Kästchen beläßt, oder aber es auf das Kästchen ein Feld links bzw. rechts von dem alten Kästchen bewegt. Außerdem geht er eventuell in einen anderen 'Geisteszustand' über. Als Konsequenz davon bleiben nur noch zwei quantitative Unbekannte übrig:
Die Zahl der zu benutzenden unterschiedlichen Zeichen und die Zahl der verschiedenen 'Geisteszustände' bzw. Berechnungsdispositionen. Wenn die Zahl der unterschiedlichen Zeichen kleiner wird, im Extremfall nur noch zwei unterschiedliche Zeichen, so wird man dementsprechend mehr unterschiedliche Berechnungsdispositionen benötigen. Reduziert man hingegen die Zahl von verschiedenen Berechnungsdispositionen, so muß zum Ausgleich die Zahl der verschiedenen Zeichen, die in ein einzelnes Kästchen geschrieben werden können, erhöht werden.
Für die formale Beschreibung eines Algorithmus werden sogenannte Turingtabellen verwendet, die die Operationen einer zugehörigen Turingmaschine beschreiben. (Siehe Abbildung 3.2.) In einer Turingtabelle wird für jeden der möglichen 'Geisteszustände' und jedes mögliche Zeichen in dem Kästchen dem das Augenmerk gilt, die drei auszuführenden Operationstypen angegeben. Durch eine solche Turingtabelle zusammen mit der Auszeichnung eines Start- und Endzustandes ist ein Algorithmus vollständig beschrieben. Ein solcher Algorithmus bzw. eine solche Turingmaschine berechnet für jede mögliche Eingabe auf dem Band, einen bestimmten Ausgabewert $A(e)$ oder aber rechnet endlos und gibt somit keinen Wert aus. Mithin kann man sagen, die Turingmaschine berechnet eine bestimmte (eventuell partielle) Funktion.

3.2 Algorithmen sind *allgemeine* Verfahren

Wie eingangs erwähnt wurde, sind Algorithmen allgemeine Verfahren, die eine beliebige Probleminstanz einer Klasse K von Problemen lösen können.[10] Bei der formalen Behandlung von Algorithmen, das heißt in der Algorithmentheorie, beinhaltet eine solche Klasse K stets abzählbar unendlich viele Probleminstanzen. Beispiele für solche Problemklassen sind:

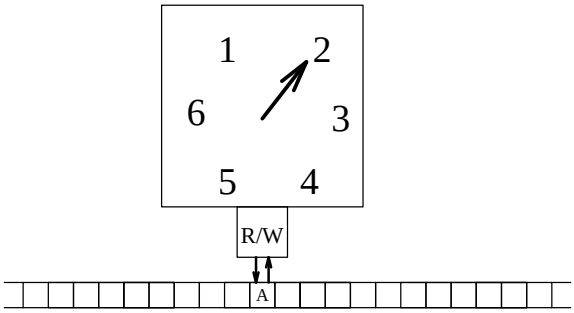
- Entscheiden, ob eine gegebene Menge von logischen Formeln widerspruchsfrei ist.
- Entscheiden, ob ein gegebener Graph eine Clique von mindestens k Knoten enthält.[11]

werden kann. Siehe beispielsweise Minsky [Min71].

[10] Man könnte auch sagen, man spricht über sogenannte μ-rekursive Funktionen, die partiell oder total auf der Menge der natürlichen Zahlen definiert sind. Partielle μ-rekursive Funktionen entsprechen bei dieser Betrachtung Algorithmen, die für manche Eingaben in Endlosschleifen geraten und damit keine Ausgabe, also kein Rechenergebnis bestimmen. Die Klasse der μ-rekursiven Funktionen ist zu der Klasse der von Turingmaschinen berechenbaren Funktionen äquivalent. Siehe für weitere Details z.B. Rogers [Rog67].

[11] Ein Graph $G = (V, E)$ besteht aus einer Menge von Knoten V und einer Menge von Kanten E, die je zwei Knoten miteinander verbinden. Eine Kante $e \in E$ ist dabei formal einfach ein Paar von Knoten aus

Turingmaschine



Turingtabelle

Zustand	x	Zeichen	→	Zustand	x	Zeichen	x	Kopfbew.
1		A	→	2		A		L
1		B	→	3		A		R
1		C	→	6		D		R
.		.	.	.		.		.
.		.	.	.		.		.
.		.	.	.		.		.
.		.	.	.		.		.
.		.	.	.		.		.
6		D	→	3		B		L

Abbildung 3.2: Das Schema einer Turingmachine und die zugehörige Turingtabelle.

Dieser Sachverhalt spielt eine ganz entscheidende Rolle für viele Aussagen, Definitionen und Theoreme in der Algorithmentheorie. Beispielsweise basiert die bekannte Unentscheidbarkeit[12] bzw. Semientscheidbarkeit[13] der Prädikatenlogik erster Stufe darauf, daß die Klasse möglicher Formelmengen, für die der gesuchte Algorithmus eine korrekte Antwort finden soll, unbeschränkt ist. Wäre die Klasse von Formelmengen, für die ein Algorithmus korrekt arbeiten soll, endlich, so gäbe es *immer* einen Algorithmus. Im Extremfall könnte man einfach ein Verfahren angeben, das die richtige Entscheidung für jede erlaubte Eingabe in einer riesigen Tabelle gespeichert hat. Dann müßte der Algorithmus lediglich in der Tabelle den passenden Eintrag aufsuchen und ihn als Antwort ausgeben. Wenn man jedoch im Gegensatz dazu beliebige Formelmengen als Probleminstanzen zuläßt, so kann man zu jeder endlichen Tabelle, die ein Algorithmus A benutzt, Formelmengen konstruieren, die der Algorithmus A nicht korrekt entscheiden kann.
Es liegt also die folgende Situation vor: Zu jeder endlichen Menge M von Probleminstanzen gibt es einen (endlichen) Algorithmus A, der jedes Problem $P \in M$ richtig löst. Hingegen gilt umgekehrt - für *un*entscheidbare Problemklassen M - daß zu jedem (end-

V. Eine Clique in einem Graphen ist eine Menge von Knoten, wobei jedes Paar von Knoten der Clique durch eine Kante miteinander verbunden sein muß.

[12] Unentscheidbarkeit bedeutet, daß es in der Prädikatenlogik für jeden möglichen Ableitungsalgorithmus A immer mindestens eine Formelmenge gibt, deren eventuelle Widersprüchlichkeit der Algorithmus A nicht korrekt entscheiden kann.

[13] Semientscheidbarkeit bedeutet, daß es einen Algorithmus gibt, der für alle Eingaben, die mit 'Ja' ('Nein') zu beantworten sind, auch die richtige Antwort findet, während er für die Eingaben, die mit 'Nein' ('Ja') zu beantworten sind, für manche Eingaben in eine Endlosschleife gerät.

lichen) Algorithmus A mindestens eine Probleminstanz $P \in M$ existiert, die A nicht korrekt lösen kann.
Um eine einheitliche Betrachtungsweise für beliebige Algorithmen zu ermöglichen, wird im folgenden der Begriff einer *universellen* Turingmaschine vorgestellt.

3.3 Universelle Turingmaschinen

Wie im ersten Abschnitt ausgeführt wurde, läßt sich jeder Algorithmus durch eine geeignete Turingmaschinentabelle beschreiben. Die Operationstabelle läßt sich ihrerseits mechanisch interpretieren. Somit liegt der Gedanke nahe, eine Turingmaschine U zu entwickeln, welche die Operationstabelle einer beliebigen Turingmaschine T interpretieren kann und damit T simuliert. Eine derartige Turingmaschine U wird auch universelle Turingmaschine genannt. Eine universelle Turingmaschine entspricht in ihrer Funktionalität heutigen Universalcomputern, die beliebig neu programmiert werden können und daraufhin ihr neues Programm ausführen. Im folgenden sollen allgemeine Betrachtungen über die Möglichkeiten von Algorithmen angestellt werden. Dafür bietet die universelle Turingmaschine den besonderen Vorteil, daß man sich auf eine konkrete Operationstabelle beschränken kann und nur noch von den verschiedenen möglichen Eingaben ausgehen muß. Somit werden alle anderen Turingmaschinen durch die Möglichkeit entsprechender Eingaben für die universelle Turingmaschine mitbetrachtet.
Eine Variante des sogenannten *Universal Turing Machine Theorems*[14] ist das folgende:

Theorem Es gibt eine universelle Turingmaschine U, die die folgende zweistellige Funktion berechnet: $U : \mathbf{N} \times \mathbf{N} \to \mathbf{N}$ wobei gilt: Sei $T(A)$ eine geeignete und berechenbare Codierung der Turingmaschinentabelle des Algorithmus A und e eine beliebige natürliche Zahl. Dann gilt $U(T(A), e) = A(e)$. U berechnet also für seine beiden Argumente (einen Algorithmus und eine Eingabe) als Ausgabewert genau den Wert, den der Algorithmus A für die Eingabe e berechnen würde.

Soll ein bestimmter Algorithmus A auf eine gegebene Eingabezeichenkette e - zum Beispiel eine Menge von Axiomen - angewendet werden, so läßt sich dies mittels einer universellen Turingmschine U also wie folgt bewerkstelligen: Man schreibt beides, die Operationstabelle $T(A)$ der Turingmaschinenrealisierung von A und die Eingabe e für A als Eingabe auf das Arbeitsband von U. Siehe Abbildung 3.3. Das Resultat der Anwendung des Algorithmus A auf die Eingabe e wird eine eventuell unendliche Zeichenfolge $r = A(e)$ sein, die U in einen festgelegten Bereich ihres Arbeitsbandes schreibt. Damit kann man also eine Beziehung zwischen der Eingabe von U, nämlich $T(A)$ verknüpft mit e und der Ausgabe r von U herstellen. Weiterhin läßt sich die Beziehung der Zahl der Eingabezeichen für U und der Länge und dem Aufbau von r herstellen. Diese Beziehung wird in der algorithmischen Informationstheorie untersucht.

[14] Siehe hierfür beispielsweise Minsky [Min71] oder Rogers [Rog67].

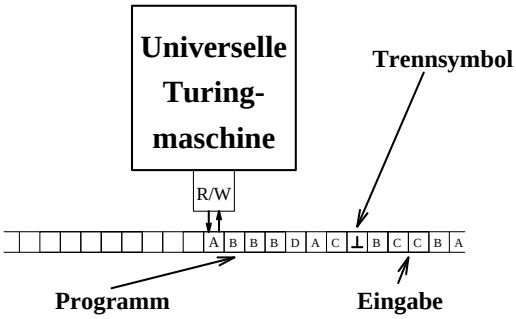


Abbildung 3.3: Die Eingabe zu einer universellen Turingmaschine.

3.4 Algorithmische Informationstheorie - Kolmogoroffkomplexität

Die algorithmische Informationstheorie ist ein kleiner und wenig bekannter Zweig der Informatik. Heute wird er im wesentlichen als Teilgebiet der Komplexitätstheorie behandelt. Entstanden ist dieses Gebiet in den sechziger Jahren durch drei voneinander unabhängige Arbeiten, die kurz hintereinander veröffentlicht wurden. Alle drei Arbeiten befaßten sich mit dem Problem der kürzestmöglichen Beschreibung einer gegebenen Zeichenkette. Die drei Autoren waren R. J. Solomonoff, A. N. Kolmogoroff und G. J. Chaitin. Solomonoff war ein Schüler Carnaps, der sich im Rahmen des Induktionsproblems damit beschäftigte, wie Beobachtungsdaten möglichst kompakt dargestellt werden können und dadurch zu induktiven Schlüssen führen. Seine Untersuchungen sind in seinem Artikel *A Formal Theory of Inductive Inference* 1964 [Sol64] erschienen. Der zweite Autor - der russische Mathematiker A.N. Kolmogoroff führte derartige Betrachtungen in seiner Arbeit *Three Approaches to the Quantitative Definition of Information* 1965 [Kol65] ein. Nach ihm wird heute auch die Länge der kürzesten Beschreibung einer Zeichenkette Z als die Kolmogoroffkomplexität von Z, $K(Z)$ bezeichnet. Chaitins Arbeit *On the Length of Programs for Computing Finite Binary Strings* [Cha66] wurde 1966 veröffentlicht.
Die minimale Länge der Beschreibung einer Zeichenkette kann von dem jeweiligen Aufbau der betrachteten universellen Turingmaschine abhängen. Daher sollen für weitere Betrachtungen die folgenden Dinge wie folgt festgelegt werden.

- Es werden nur *binäre* endliche oder unendliche Zeichenketten betrachtet. Dies gilt sowohl für Zeichenketten, die beschrieben werden sollen, wie für beschreibende Zeichenketten. Die beschreibenden Zeichenketten - also die jeweiligen Eingaben für eine universelle Turingmaschine U - werden auch als *Programm* bezeichnet.

- Es werden nur universelle Turingmaschinen betrachtet, deren vollständige Turingmaschinentabelle nicht mehr als 1000 Zeilen enthält.[15]

[15] Ein Beispiel für eine entsprechende universelle Turingmaschine ist in Minsky [Min71] auf Seite 190 angegeben.

- Das kürzeste Programm für eine universelle Turingmaschine U, um eine Zeichenkette Z zu konstruieren, wird im folgenden auch als Kolmogoroffkomplexität von Z, kurz $K(Z)$ bezeichnet.

Ein grundlegendes Theorem der algorithmischen Informationstheorie ist das sogenannte *Invariance Theorem* welches besagt, daß der besondere Aufbau der betrachteten universellen Turingmaschine U lediglich einen Unterschied in der Kolmogoroffkomplexität einer Zeichenkette Z von einer Konstanten c ausmacht. Dies läßt sich einsehen, indem man sich überlegt, daß auf jeder universellen Turingmaschine U jede andere universelle Turingmaschine U' durch ein Programm von fester Länge c simuliert werden kann. Mithin ist c als eine kleine Zahl anzunehmen, wenn man von überschaubaren Varianten einer universellen Turingmaschine ausgeht (ihre Beschreibung darf nicht zu lang werden). Das Invariance Theorem ist geradezu Voraussetzung für die algorithmische Informationstheorie, da ansonsten alle Aussagen stets auf eine spezifische universelle Turingmaschine bezogen werden müßten. Im Gegensatz dazu kann man aufgrund des Invariance Theorems von der Komplexität sprechen, die den Zeichenketten inhärent ist und nicht von einem Interpretationsmechanismus abhängt.
Das Invariance Theorem ist die theoretische Grundlage dafür, daß man von der Komplexität von Zeichenketten in einem gewissen *absoluten* Sinn sprechen kann.
Diese Tatsache der absoluten Komplexität von Zeichenketten, das heißt, der Unmöglichkeit Zeichenketten unter eine bestimmte Mindestlänge zu komprimieren - sie nicht kürzer beschreiben zu können, soll Gegenstand der nachfolgenden Kapitel sein.

Einige kombinatorische Überlegungen

Es gibt 2^n verschiedene binäre Zeichenketten der Länge n. Da jede Zeichenkette zu ihrer Konstruktion ein anderes Programm erfordert, gibt es stets einige Zeichenketten der Länge n, deren Kolmogoroffkomplexität mindestens n ist. Weiterhin läßt sich zeigen, daß sogar die allermeisten Zeichenketten der Länge n eine Kolmogoroffkomplexität der Größenordnung n haben.
Interessanterweise läßt sich erkennen, daß es stets eine Zeichenkette Z der Länge n gibt, wobei $K(Z)$ mindestens n ist, während sich ein solches Z *nicht* für den *allgemeinen* Fall konkret angeben läßt - auch keine allgemeine Konstruktionsvorschrift ! Dies rührt daher, daß die kürzeste Beschreibung von Z nach Definition gerade mindestens von der Länge n ist, mithin es kein allgemeines Schema von konstanter Länge geben kann, nach dem sich Z für ein beliebiges n bestimmen läßt.
Dieser Sachverhalt läßt bestimmte Parallelen zu dem Gödelschen Theorem[16] über die Unvollständigkeit von Axiomensystemen für die elementaren Arithmetik erkennen [Göd31]. Dort zeigte Gödel, daß es für jedes mögliche, widerspruchsfreie Axiomensystem A stets wahre Sätze der elementaren Arithmetik gibt, die sich aus A nicht ableiten lassen. In der Tat gelang es Chaitin, das Gödelsche Unvollständigkeitstheorem mit Hilfe der algorithmischen Informationstheorie zu beweisen [Cha74].

[16] Eine ausführliche Darstellung von Gödels Beweis ist in Nagel & Newman [NN64] zu finden.

Beispiele: Zeichenketten wie '11111111', '000000000000000' oder '101010101010101010' etc. sind einfache Zeichenketten, die sich mittels eines kurzen Programms konstruieren lassen. Ein derartiges Programm muß im wesentlichen nur die Länge der Zeichenkette beinhalten; also höchstens $\log_2$ (Länge(Z)) Binärzeichen lang sein, sowie noch ein bis zwei weitere Zeichen beinhalten, die bestimmen, ob Einsen oder Nullen ausgedruckt werden sollen. Je nach dem wie lang eine Zeichenkette ist, läßt sich ihre Länge auch noch kürzer darstellen, beispielsweise wie bei den folgenden Zahlen: 10^{10}, $9^{(9^9)}$, etc.
Dagegen sind Zeichenketten wie '10000111101001010101011101011010111011110000010...' oder '10011010010010110111110110101110101001 0...' komplizierter, d.h. sie haben eine größere Kolmogoroffkomplexität, da zur Beschreibung solcher Zeichenketten die Angabe ihrer Länge bei weitem nicht ausreicht.

Für die folgenden Ausführungen ist bemerkenswert, daß Zeichenketten, insbesondere *endliche* Zeichenketten auf sehr verschiedene Weisen algorithmisch beschrieben werden können.

0 1 0 0 0 1 1 0 1 1 1 1 0 0 0 1 1 0 0 ...
Dies kann wie folgt beschrieben werden:

1. Es stehen überall Nullen, außer an den Stellen 2, 6, 7, 9, 10, 11, 12, 16, 17, ...
2. Es stehen alternierend jeweils 2 Einsen und 3 Nullen; Ausnahmen sind die Stellen 1, 9, 10, ...

Anwendungen und Entwicklungen der Kolmogoroffkomplexität

Die Anwendungen des Begriffs der Kolmogoroffkomplexität liegen primär in den folgenden Gebieten:

- Anwendung der möglichen Komprimierung von Zeichenketten: Zu den berühmtesten Anwendungen gehört sicherlich Chaitins Version des Beweises des Gödelschen Unvollständigkeitstheorems in Chaitin [Cha74]. Eine andere Anwendung der Komprimierungsbetrachtungen ist das induktive Schließen. Dort wird versucht, eine möglichst kurze Beschreibung für die gegebenen empirischen Daten zu finden. Siehe beispielsweise Solomonoff [Sol64], Freivalds & Kinber [FK77] oder Li & Vitanyi [LV89].
- Anwendungen in der mathematischen Zahlentheorie: Beispielsweise untersuchte Chaitin [Cha87] die Lösbarkeit diophantischer Gleichungen[17] und zeigte dabei, daß die Lösung bestimmter diophantischer Gleichungen einen unbegrenzten algorithmischen Informationsgehalt erfordert.

[17] Diophantische Gleichungen sind Gleichungen, bei denen auf beiden Seiten des Gleichheitszeichens arithmetische Ausdrücke stehen, die nur die Operationen der Addition, der Multiplikation und der Potenzierung beinhalten. Dabei sind einige Zahlenwerte durch Konstanten vorgegeben, während andere Variablen sind. Für die Variablen muß eine Belegung gefunden werden, die die Gleichung erfüllt. Ein einfaches Beispiel ist das folgende: $3^n + 4^n = 5^n$. Die Gleichung gilt für $n = 2$.

- Die Anwendung der Tatsache, daß es Zeichenketten gibt, die nicht kompromiert werden können: Die nicht komprimierbaren Zeichenketten lassen sich nicht effektiv konstruieren, obwohl es durch das erwähnte Zählargument offensichtlich ist, daß es solche Zeichenketten gibt. Diese Tatsache kann genutzt werden, um elegante und einfache Beweise über Unterschranken von Berechnungskomplexitäten zu führen. Siehe beispielsweise Paul [Pau79]. Weitere Anwendungen sind von Li & Vitanyi in [LV91] diskutiert oder in Kirchherr [Kir92] zu finden.

Eine Weiterentwicklung des vorgestellten Begriffs der Kolmogoroffkomplexität stellt die *resourcenbeschränkte Kolmogoroffkomplexität* dar: Hier werden die erlaubten Resourcen[18] beschränkt, die zur Erzeugung einer Zeichenkette mittels der komprimierten Zeichenkette (dem Programm für die universelle Turingmaschine) erforderlich sind. Dies nutzte beispielsweise Adleman [Adl79] um die Zeitkomplexität des Faktorisierungsproblems[19] zu untersuchen.
In der Arbeit werden nun mit Hilfe des Begriffs der allgemeinen Kolmogoroffkomplexität eine Reihe von Betrachtungen über die Natur künstlicher Intelligenz und kognitiver Prozesse angestellt. Die Kolmogorffkomplexität dient dabei als Maß der Komplexität einer (statischen) Beschreibung dieser Prozesse - sie hat also noch nichts mit der eventuellen Ausführung eines Algorithmus und damit etwa mit der Zahl der dabei benötigten Rechenschritte zu tun.
Will man jedoch konkreter werden, und bestimmte Probleme im Detail betrachten, so bietet es sich an, dabei auch einen Begriff der resourcenbeschränkten Kolmogoroffkomplexität zu verwenden. Denn einerseits kann man einen Computer bei der Ausführung eines Programms nicht unbeschränkt viel Zeit einräumen und andererseits sind auch bestimmte kognitive Modelle nicht plausibel, wenn man davon ausgeht, daß das menschliche Gehirn als 'biologische Rechenmaschine' unbeschränkte Resourcen zur Verfügung hat.[20]

Zur Charakterisierung von Chaos

Es sei angemerkt, daß der Begriff des *Chaos* [21] mit Hilfe des Begriffs der Kolmogoroffkomplexität eine scharfe Fassung erhalten kann. Wenn man gemeinhin *Chaos* als 'Regellosigkeit' versteht, so kann eingewendet werden, daß die 'Regellosigkeit' gerade selbst wieder eine Regel ist, mithin diese Auffassung des Chaosbegriffs ad absurdum geführt ist.

[18] D.h. Zahl der Rechenschritte und/oder der Speicherplatz sind beschränkt.

[19] Faktorisierung nennt man das Problem, das Produkt zweier Primzahlen in seine beiden Faktoren zu zerlegen. Dieses Problem hat Anwendungen in der Kryptographie. Die Faktorisierung einer bekannten großen Zahl (mehrere Hundert Stellen) ist dabei gleichbedeutend mit der Entschlüsselung der Nachrichten.

[20] Beispielsweise sieht Valiant [Val84a] komplexitätstheoretische Schranken nicht nur als unüberwindlich für Computer an, sondern auch für das menschliche Denken.

[21] Briggs bietet eine allgemeinverständliche Einführung in die physikalische Chaosforschung [BP90]. Dort wird auf physikalische Modellbildungen hingewiesen, für die unter Umständen keine adäquaten Berechnungsalgorithmen möglich sind. Dies bedeutet allerdings nicht, daß man die durch die Modelle beschriebenen Phänomene überhaupt nicht algorithmisch beschreiben kann. Schließlich läßt sich für alles Endliche, was überhaupt geschieht und beschreibbar ist, auch eine *algorithmische* Beschreibung angeben. Die Frage, wie man zu einer vorhersagenden Phänomenbeschreibung gelangen kann, ist hiervon noch nicht berührt.

Mit dem oben definierten Begriff der *Kolmogoroffkomplexität* läßt sich Chaos als die Tatsache auffassen, daß die Beschreibung der chaotischen Phänomene eine große Kolmogoroffkomplexität erfordert. Hierbei würde die 'Regellosigkeit' aus der ersten Begriffsbestimmung durch eine große (z.B. 10^{10} Bits) Kolmogoroffkomplexität vertreten. Der Einwand, daß dies selbst wieder eine Regel sei, würde sich hier deutlich als eine Aussage über die *Beschreibung* des Chaos, also als eine Metaaussage herausstellen. Die 'Regellosigkeit' ist somit keine Eigenschaft der Phänomene, sondern eine Eigenschaft ihrer Beschreibung.[22]

[22] Siehe beispielsweise die Arbeit von S. Ambroskiewicz [Amb91] der den Zusammenhang zwischen Chaos, dessen hohe Kolmogoroffkomplexität und der damit implizierten probabilistischen Zusammenhänge untersucht.

Kapitel 4

Symbolmanipulation und Intentionalität

In diesem Kapitel soll die Beziehung von Intentionalität, d. h. von Meinungen, Wünschen, etc. und formaler Symbolmanipulation erörtert werden. Im ersten Abschnitt wird Pylyshyns Konzeption einer Fundierung der Kognitionswissenschaft dargestellt. Pylyshyn argumentiert dafür, daß Kognitionen im *buchstäblichen* Sinn Symbolmanipulationen sind. Daß sich die menschlichen Intelligenzleistungen *überhaupt* durch Symbolmanipulation erklären und beschreiben lassen, ist allerdings umstritten. Im zweiten Abschnitt werden Kritikpunkte an dieser Hypothese, der sogenannten *Wissensrepräsentationshypothese*, die auch Pylyshyns Überlegungen zugrunde liegt, erörtert.

4.1 Über eine Grundlegung der Kognitionswissenschaft

Der Gegenstand der Kognitionswissenschaft ist nicht so klar, wie er zunächst vielleicht erscheinen mag. Die Natur des menschlichen Denkens soll erforscht werden. Doch wie dies zu geschehen hat und wie mögliche Forschungsergebnisse aussehen müssten oder könnten, ist damit noch nicht geklärt. Pylyshyn versuchte eine Grundlegung der Kognitionswissenschaft in seinem Buch 'Computation and Cognition', 1984 [Pyl84]. Die Hauptthese, die Pylyshyn in seinem Buch vertritt ist, daß in einer sinnvollen empirischen Theorie der Kognitionen - des Denkens - drei verschiedene Ebenen der Beschreibung zu unterscheiden sind. Jede der drei Ebenen hat ihre eigenen Prinzipien nach denen Prozesse auf ihnen zeitlich parallel ablaufen. Dabei gibt es spezifische Abhängigkeiten zwischen den drei Prozeßebenen. Pylyshyn unterscheidet im einzelnen die *biologische* bzw. *physische Ebene*, die *symbolische* bzw. *syntaktische* oder *funktionale Ebene* sowie die *semantische* oder *intentionale Ebene*. Beispielsweise bestimmen biologische Faktoren in welcher Geschwindigkeit bestimmte 'Denkoperationen' auf der symbolischen Ebene durchgeführt werden. Prozesse auf der symbolischen Ebene sind ihrerseits wesentlich dafür verantwortlich, inwiefern sich ein Individuum rational verhält bzw., ob auf der semantischen Ebene tatsächlich die logi-

schen Schlüsse gezogen werden, die auf Grund des vorhandenen *Wissens*[1] gezogen werden könnten. Auf diese Art und Weise könnte im Prinzip auch eine detaillierte Theorie der Denkprozesse in besonderen psychischen Zuständen entwickelt werden, die beispielsweise von Halluzinationen bestimmter Art begleitet werden, wobei solche Halluzinationen möglicherweise durch abnorme Vorgänge auf der biologischen Ebene charakterisiert werden könnten. Solch abnorme Vorgänge hätten entsprechend veränderte Prozesse auf der symbolischen und auf der semantischen Ebene zur Folge. Somit könnte man Halluzinationen auf der semantischen Ebene als Folge von veränderten Prozessen auf der biologischen Ebene erklären.
Häufig ist man in der künstlichen Intelligenz versucht, sich nur auf die syntaktische und semantische Ebene zu konzentrieren. Dies mag für die Erstellung von funktionierenden intelligenten Computerprogrammen auch in der Regel ausreichen; um allerdings auf diese Weise Theorien zu finden, die menschliches Verhalten erklären, erscheint Pylyshyn zu kurz gegriffen. Um menschliches Verhalten vollständig zu erklären, wird man ohne die biologische Ebene nicht auskommen, deren Prozesse die spezifischen Prozesse auf der syntaktischen Ebene bestimmen. Wenn die Prozesse auf der semantischen Ebene vollständig von den Prozessen auf der syntaktischen Ebene abhängen, und die Prozesse auf der syntaktischen Ebene ihrerseits wiederum vollständig von den Prozessen auf der biologischen Ebene abhängen, könnte man sich im Prinzip darauf beschränken, nur die Prozesse auf der biologischen Ebene zu beschreiben. Alle anderen Prozesse wären implizit in der Beschreibung der Prozesse auf der biologischen Ebene enthalten. Ein solch reduktionistischer Ansatz würde allerdings dazu führen, daß die Beschreibung von Denkprozessen auf der biologischen Ebene ungeheuer kompliziert wäre. Es wäre praktisch unmöglich, daraus die Konsequenzen für die semantische Ebene von Denkprozessen, also beispielsweise für konkrete Gedankeninhalte, zu ziehen.[2] Ein Grund dafür, daß Pylyshyn gerade die genannten drei Ebenen unterscheidet, ist die Annahme, daß die Prozesse auf den drei Ebenen nach unterschiedlichen Prinzipien ablaufen. Ein zweiter, vielleicht noch wichtigerer Grund für diese Unterscheidung ist die Tatsache, daß bereits mehr oder weniger ausgearbeitete Theorien über die Prozesse auf diesen Ebenen vorliegen. Auf der biologischen Ebene gibt es bereits wissenschaftliche Theorien über biochemische bzw. neurophysiologische Phänomene, die ihre Wurzeln in den traditionellen Naturwissenschaften haben. Für die syntaktische oder funktionale Ebene sind aus den Überlegungen zur künstlichen Intelligenz und aus verschiedenen Ansätzen zu Gedächtnistheorien ebenfalls schon theoretische Vorarbeiten geleistet worden.[3]

[1] Wissen zählt auch zur semantischen Ebene.

[2] Die größere Komplexität auf der biologischen Ebene bezieht sich hier in erster Linie auf die Betrachtung der dynamischen Aspekte. Gemessen in Kolmogoroffkomplexität kann ein Unterschied in der Komplexität in der statischen Beschreibung daher rühren, daß auf der biologischen Ebene viele 'technische' Details beschrieben werden müssen, die von keinerlei Relevanz für die syntaktische und semantische Ebene sind. Dies ist allerdings nur denkbar, wenn sich die biologischen Details in ihrer Funktion sozusagen gegenseitig auslöschen. (Etwa wie es bei sich überlagernden Wellen in der Physik der Fall ist.)

[3] Stich [Sti83] argumentiert dafür, daß die syntaktische Ebene für die Kognitionswissenschaft gegenüber der semantischen Ebene zu bevorzugen sei. Stich weist unter anderem darauf hin, daß die syntaktische Ebene in jedem Fall ausreicht, während mit der semantischen Ebene viele unnötige Schwierigkeiten ver-

Für die dritte, die semantische oder intentionale Ebene sind aus dem alltäglichen Denken und Reden über mentale Phänomene, sowie aus introspektiven Einsichten heraus zumindest rudimentäre Ansätze zu Theorien vorhanden. Die genannten Theorien bzw. Ansätze zu Theorien beinhalten ontologische Annahmen, die die vorgenommene Dreiteilung nahelegen.
Eine von Pylyshyns wichtigsten empirischen Behauptungen ist, daß es einen 'natürlichen' Bereich von Phänomenen gibt, den man als einen privilegierten Bereich für eine algorithmische Ebene ansehen kann.[4] Später werden noch Bedingungen formuliert werden, um auf dieser Basis angemessene erklärende Theorien zu entwickeln.
Anschließend sollen die grundlegenden Kriterien herausgearbeitet werden, die Verhalten, das durch Regeln und Repräsentationen gesteuert wird, von Verhalten unterscheiden, das bloß das Ergebnis der kausalen Struktur des zugrunde liegenden biologischen Systems ist. Ein zentrales Ziel von Pylyshyns Arbeit liegt darin, einen solchen Unterschied als *prinzipiell* aufzuzeigen und dabei ein *notwendiges* empirisches Unterscheidungskriterium anzugeben. Dieses methodologische Kriterium, das Pylyshyn *cognitive penetrability condition* nennt, unterstützt die These, daß eine erklärende Theorie eine Art von Symbolmanipulationsmodell sein sollte.

4.1.1 Intentionen und formale Symbolverarbeitung

Nach der von Pylyshyn vorgeschlagenen Trennung einer semantischen Ebene von einer syntaktischen oder funktionalen Ebene in der Kognitionswissenschaft, wäre *Denken als Symbolverarbeitung* zu begreifen. Der Gedanke, daß Kognitionen als regelgeleitetes Manipulieren von Symbolen aufgefaßt werden können, also Denken als Informationsverarbeitung verstanden werden kann, steht nun mittlerweile seit mehr als zwei Jahrzehnten im geistigen Umfeld der Kognitionswissenschaft.[5]
Zunächst soll mit einer informellen Diskussion der Behauptung begonnen werden, daß bestimmte Arten des menschlichen Verhaltens durch *Repräsentationen* (z.B. von Meinungen, Zielen,...) bestimmt werden. Hierbei ist gemeint, daß Symbole, die auf eine unbestimmte Art physikalisch realisiert sind, Meinungen, Ziele, Wünsche, Hoffnungen, usw. repräsentieren. Mithin unterscheidbare Symbole stellvertretend für unterscheidbare Meinungen, Ziele, usw. stehen. Aus diesem Grund werden häufig mentale Prozesse als Symbolmanipulationsvorgänge betrachtet. Zunächst soll auf mögliche Vorteile einer repräsentationalen Beschreibungsebene hingewiesen werden.
Die meisten Menschen besitzen eine implizite, ausgeprägte und sehr erfolgreiche kognitive Theorie. Sie können dadurch einen enormen Bereich menschlichen Verhaltens systematisieren und korrekte Vorhersagen machen. Obwohl es genug Fehler und Irrtümer in dieser Alltagspsychologie gibt, übertrifft sie bei weitem die gegenwärtige wissenschaftliche Psychologie sowohl in ihrem Anwendungsbereich als auch in ihrer Präzision. Das Interessante

bunden sind. Eine Diskussion darüber findet man auch in Block [Blo90].

[4] Vergleiche hierzu die kritische Diskussion zu den Prinzipien einer künstlichen Intelligenz in Kapitel 6.

[5] Siehe z.B. auch Haugeland [Hau78], Fodor [Fod87a], Cummins [Cum89] oder Block [Blo90] für einen Überblick über die Dikussion.

dabei ist das mentalistische Vokabular der Alltagspsychologie sowie die zugrundegelegte Taxonomie von Dingen, Verhalten, Ereignissen usw.. Die Abstraktionen, die sich in den Begriffen der Alltagspsychologie zeigen, und insbesondere deren repräsentationaler Charakter (z.B. subjektive Meinungen, Ziele, Interpretationen, etc.) erscheint außerordentlich gut geeignet, um die Generalisierungen aus dem Bereich der kognitiven Psychologie zu erfassen. Es scheint, daß kognitive Phänomene durch Meinungen, Absichten etc. erklärt werden müssen, da bestimmte Regelmäßigkeiten menschlichen Verhaltens nur durch solche Begriffe erfaßt werden können.
Wenn beispielsweise eine Person Gefahr verspürt, wird sie versuchen sich von der Gefahrenquelle zu entfernen. Diese Generalisierung hat eine unüberschaubare Zahl von Instanzen. Angenommen eine Person weiß allgemein, wie sie sich aus einem Gebäude entfernen kann. Wenn sie glaubt, daß das Gebäude brennt, in dem sie sich befindet, so wird sie sich das allgemeine Ziel setzen, das Gebäude zu verlassen. Dabei wird sie ihr Wissen dazu verwenden, die einzelnen Schritte dieser Handlung zu bestimmen und auszuführen. Eine solch einfache Regularität wird sich kaum ohne mentalistische Begriffe wie *meinen, glauben,* etc. beschreiben lassen.
Denn es gibt eine Vielzahl von Möglichkeiten zu wissen, wie man ein Gebäude verläßt oder zu glauben, daß es brennt. Jede dieser unterschiedlichen Möglichkeiten würde auf einer anderen Beschreibungsebene, z.B. auf einer neurophysiologischen Ebene, vollkommen unterschiedliche kausale Ketten von Ereignissen hervorrufen. Damit würden jedoch die psychologisch relevanten Generalisierungen in der Vielzahl von unterschiedlichen Ereignisketten verloren gehen. Es ist sicher kein Zufall, daß die Systematisierung des menschlichen Verhaltens durch ein Vokabular geschieht, das auf interne Repräsentationen von Meinungen, Absichten, Zielen usw. referiert.
Wenn jemand beispielsweise einen Text schreibt, so hat er dabei die Absicht einige Aussagen zu machen und nicht etwa irgendwelche Tintenkleckse zu fabrizieren, die allerdings bei der Handlung auch noch entstehen. Eine Theorie, die die Fingerbewegungen in den Vordergrund stellt, könnte nicht erklären, warum jemand plötzlich ganz andere Bewegungen macht, wenn seine Schreibmaschine kaputt ist und er deswegen zu Papier und Bleistift greift. Die Alltagspsychologie hat natürlich auch einige Schwächen, wenn man an sie den Anspruch stellt, eine wissenschaftliche Theorie zu sein. Es mag hier eine Menge von überflüssigen oder bedeutungslosen Begriffen geben und viele Erklärungen mögen auch unvollständig oder zirkulär sein. Die schwerwiegendste Schwäche scheint jedoch die Tatsache zu sein, daß die Menge von losen Generalisierungen die dieses informale Wissen ausmacht, nicht zu einem expliziten System verbunden ist, woraus systematisch Aussagen abgeleitet werden könnten.

4.1.2 Denken als Symbolverarbeitung

Pylyshyn argumentiert dafür, daß Denken tatsächlich als Symbolverarbeitung betrachtet werden sollte, und zwar in einem buchstäblichen Sinn ! Er begründet dies damit, daß Symbolverarbeitung die einzige ausgearbeitete Beschreibung eines Prozesses ist, die sich mit zwei verschiedenen Beschreibungsebenen vereinbaren läßt; mit der materialistischen Sicht

wie ein Prozeß physikalisch realisiert ist einerseits, und andererseits mit der Beschreibung des Verhaltens eines Prozesses durch regelgeleitete Operationen auf Symbolen.
Symbolverarbeitung als auch Denken sind Prozesse von einem grundlegend gleichen Typ, da sie beide physikalisch realisiert sind, aber trotzdem durch Regeln und Repräsentationen gesteuert werden. Da ein Computer ein physikalisches Gerät ist, kann das Verhalten eines Computers durch die Beschreibung der kausalen Struktur seiner physikalischen Eigenschaften beschrieben werden. Die (physikalischen) Zustände eines Computers sind durch Begriffe der physikalischen Beschreibung unterschieden und seine Zustandsübergänge sind deswegen durch physikalische Gesetze bestimmt. Wenn von diesen physikalischen Eigenschaften abstrahiert wird, so läßt sich eine funktionale Beschreibung des Gerätes angeben. Beispielsweise geht ein Computer immer wenn er in einem Zustand m ist, in einen Zustand n über. Dies ist kein physikalisches Gesetz, obwohl dieser bestimmte Computer gerade so aufgebaut ist, daß er kraft physikalischer Gesetze, den besonderen physikalischen Eigenschaften, die er aufweist, wenn er im Zustand m ist sowie seines spezifischen Aufbaus *immer* von dem Zustand m in den Zustand n übergeht.
So kann solch eine funktionale Beschreibung eines Computers durch ein geeignetes Zustandsübergangsdiagramm geschehen. Dies ist allerdings in der Regel keine angemessene Beschreibung, um die funktionale Struktur des Computers *leicht* verstehen zu können.
Wenn erklärt werden soll, was für eine Berechnung ein Computer ausführt, bzw. welche Regularitäten ein bestimmter, besonders programmierter Computer aufweist, so muß man auf die Objekte desjenigen Bereiches verweisen, in den die *intendierte* Interpretation der Berechnungen des Computers fällt. Dies könnte beispielsweise der Bereich der ganzen Zahlen sein, wenn die Symbolmanipulationen des Computers als numerische Berechnungen interpretiert werden sollen. Um zu erklären, warum der Computer das Zeichen '5' als Reaktion auf die Eingabe 'Add 2, 3' ausdruckt, müssen die Zeichen *Add* als Additionsoperator in den natürlichen Zahlen, sowie die Zeichen '2' und '3' als dessen Operanden interpretiert werden. Dadurch kann die Reaktion des Computers als die Auswertung der Additionsoperation auf den eingegebenen Operanden erklärt werden.
Was Computer tun, wird typischerweise erklärt, indem auf eine bestimmte intendierte Interpretation referiert wird. Allerdings erklärt man auf die gleiche Art, warum und wie Menschen das tun, was sie tun. Bei der Erklärung warum ein Schachspieler eine bestimmte Figur auf ein bestimmtes Feld setzt, wird dies durch die Rolle der Figur, die sie im Schachspiel spielt sowie die besonderen Ziele und Pläne beschrieben, die der betreffende Schachspieler in der vorliegenden Situation verfolgt. Man versucht nicht, die Handlung des Schachspielers dadurch zu erklären, daß man auf den spezifischen neurophysiologischen Zustand seines Gehirns eingeht oder die geometrischen Eigenschaften der bewegten Schachfigur beschreibt. Man geht auf eine geeignete Interpretation der Spielsituation ein.

Pylyshyn sieht hier einen fundamentalen Unterschied zwischen der Beschreibung dessen, was ein Computer tut, durch die Beschreibung seiner internen Zustände (z.B. durch Äquivalenzklassen physikalischer Beschreibungen) und der Beschreibung durch das, worauf sich das Tun des Computers bezieht. Den fundamentalen Unterschied sieht er darin, daß die erste Art von Beschreibung auf *intrinsische* Eigenschaften des Gerätes referiert,

während die zweite Art von Beschreibung auf einen vollkommen anderen Bereich referiert, der in der Regel nichts mit dem physikalischen Aufbau des Gerätes zu tun hat.
Diesem fundamentalen Unterschied bei der Beschreibung von Computern oder anderen ähnlich komplexen Geräten, entspricht bei der Beschreibung des menschlichen Denkens auf der einen Seite die *neurophysiologische* oder kausale Beschreibung und auf der anderen Seite die *intentionale* Beschreibung. Dieser Unterschied hat der Philosophie lange Zeit schwere Probleme bereitet. Das Problem entsteht durch die Frage, wie es möglich ist, daß Menschen ein bestimmtes Verhalten aufgrund von bestimmten Zielen, Wünschen und Annahmen zeigen, während man gleichzeitig davon ausgeht, daß sie aufgrund von physikalischen Gesetzmäßigkeiten und ihrem besonderen neurophysiologischen Zustand sich so verhalten.[6] Man kann allerdings mit der gleichen Verwunderung die entsprechende Frage bei dem Verhalten von Computern stellen: Wie können die Zustandsübergänge einer Maschine gleichzeitig von physikalischen Gesetzmäßigkeiten und von den abstrakten Eigenschaften der ganzen Zahlen abhängen ?
Die Antwort ist die Folgende: Dies ist dadurch möglich, daß beides, sowohl die Zahlen als auch die Regeln, die die Beziehungen zwischen den Zahlen beschreiben, in der Maschine als symbolische Ausdrücke und Programme repräsentiert sind. Die physikalische Realisierung dieser Repräsentationen schließlich ist dafür verantwortlich, daß sich die Maschine tatsächlich gemäß der repräsentierten Regeln verhält. Computer wenden nur aufgrund der formalen Eigenschaften der repräsentierten Ausdrücke Regeln an. Mithin können sie nur rein syntaktisch arbeiten und können hierbei keinerlei Unterschiede zwischen der Interpretation der Ausdrücke und den Ausdrücken selbst machen. Daher ist es zwingend notwendig, daß alle relevanten Eigenschaften des Interpretationsbereiches auch eine Entsprechung in den verwendeten Ausdrücken finden. Kurz, alle semantischen Unterschiede müssen sich auf der syntaktischen Ebene der verwendeten Ausdrücke wiederfinden.
Durch die Trennung der semantischen und der syntaktischen Aspekte kognitiver Prozesse wird das Problem der intentionalen Handlungen darauf reduziert, einen geeigneten Mechanismus anzugeben, der auf bedeutungsfreien Symbolen - auf Zeichen - operiert und dabei bedeutungsvolle Prozesse - zum Beispiel Rechnen - ausführt. Dies heißt aber auch, daß der Computer vollkommen unabhängig von der Interpretation seiner verwendeten Symbole arbeitet. Insofern kann man sagen, daß der Computer nicht weiß, was er tut, - wie beispielsweise Fodor [Fod78] oder Dreyfus [Dre72]. Die syntaktische repräsentationsgesteuerte Natur von Computerberechnungen zeigt somit wie eine Beziehung von Symbolmanipulationen zu Kausalgesetzen - zumindest im Prinzip - herstellbar ist.
Darüberhinaus erfordert die Erklärung kognitiver Gesetzmäßigkeiten nicht mehr notwendig auch die Erklärung der neurophysiologischen Vorgänge im Gehirn. Zumindest ist dies nicht notwendiger, als die Erklärung der physikalischen Realisierung eines Computers bei der Erklärung von Computerberechnungen. Allerdings ist für beide Erklärungen gleichermaßen wesentlich, daß die elementaren Operationen des zugrunde liegenden Systems sowie die Steuerung der Ausführung der elementaren Operationen spezifiziert werden.

[6] Fodor [Fod87a] hat hierbei eine Kausaltheorie der Repräsentation vorgeschlagen. Cram [Cra92] diskutiert Probleme Fodors Theorie und versucht sie mit Dretskes Idee [Dre86] für mögliche Mißrepräsentationen zu vereinbaren.

Diese geforderte Spezifikation bezeichnet Pylyshyn auch als die *funktionale Architektur* des symbolmanipulierenden Systems. In diesem Sinn sieht Pylyshyn Symbolmanipulationen nicht nur als *metaphorische*, sondern als *buchstäbliche* Beschreibungsmöglichkeit von Denken an.[7]

4.1.3 Funktionale Architektur und kognitive Prozesse

Bei der Simulation von kognitiven Prozessen fordert Pylyshyn eine *starke Äquivalenz*. Es soll nicht nur das gleiche Ergebnis bei der Simulation herauskommen, sondern die einzelnen Simulationsschritte sollen zu den tatsächlichen Schritten des zu simulierenden kognitiven Prozesses korrespondieren. Damit werden an die Computersimulation bei kognitiven Prozessen wesentlich höhere Ansprüche als sonst gestellt. Ein adäquates Computermodell muß die *funktionale Architektur*, auf dem der kognitive Prozeß abläuft, widerspiegeln, um darauf aufbauend den repräsentationalen Charakter von Meinungen, Wünschen, Zielen usw., die für den kognitiven Prozeß wesentlich sind, entsprechend darstellen zu können.
Was ein einzelner Schritt eines kognitiven Prozesses ist, hängt damit von der funktionalen Architektur ab, die die einzelnen Elementaroperationen zur Verfügung stellt. Dies ist mit den einzelnen elementaren Befehlen einer Programmiersprache zu vergleichen. Die besagte funktionale Architektur ist bei Computern in der Regel elektrisch realisiert, bzw. sie wird durch elementare kleine Programme auf niedriger Programmierebene (Maschinenebene) realisiert. Eine solche funktionale Architektur, die nicht direkt elektrisch realisiert ist, sondern durch kleine elementare Programme nur die äußere Form einer unmittelbar physikalisch (, z.B. elektrisch) realisierten funktionalen Architektur hat, wird auch *virtuelle Maschine* genannt.
Pylyshyn geht davon aus, daß beim Menschen die elementaren Operationen seiner (kognitiven) funktionalen Architektur durch den neurophysiologischen Aufbau seines Gehirns bereitgestellt werden. Inwiefern diese elementaren Operationen ihrerseits wiederum durch einen *subkognitiven Prozeß* beschrieben werden können, soll uns an dieser Stelle nicht interessieren.
Die Computersimulation soll also den gleichen Algorithmus ausführen, der bei dem menschlichen, *kognitiven* Prozeß abläuft.[8] Das heißt, der Startzustand repräsentiert eine bestimmte Menge von Meinungen, Zielen, etc. und jeder Zwischenschritt bis zum Erreichen des Endzustands des Algorithmus repräsentiert wiederum eine Menge von Meinungen usw., die ihre Entsprechung in einem Zwischenschritt des tatsächlich ablaufenden kognitiven Prozesses finden.
Bei diesem Anspruch, dessen Erfüllung Pylyshyn als das Ziel der Kognitionswissenschaft ansieht, tritt allerdings das folgende grundsätzliche Problem auf: Woher weiß man - außer durch wenig fundierte methodologische Annahmen - wie die funktionale Architektur aussieht, und welcher Art die darauf basierenden Regeln und die Repräsentationen von

[7] Dies steht beispielsweise im Gegensatz zu Arbib [Arb89], der lediglich eine metaphorische Korrespondenz sieht.

[8] Daß die menschlichen kognitiven Prozesse algorithmisch sind, wird allerdings u.a. von Penrose [Pen89] bestritten. Hierauf wird in Abschnitt 9.5 noch näher eingegangen.

Meinungen, Zielen usw. sind ?
Wie die funktionale Architektur aussieht, ist allerdings nicht das einzige Problem. Bei der Verwendung von Symbolverarbeitungsmodellen als Erklärungen von kognitiven Prozessen ist noch ein weiteres schweres Problem zu lösen:

> Es ist eine Menge von Phänomenen zu isolieren, die das Symbolverarbeitungsmodell konstituieren und mittels derer die kognitiven Prozesse erklärt werden können. Dabei ist zu unterscheiden, wie die funktionale Architektur einerseits und der symbolverarbeitende Prozeß *auf* der funktionalen Architektur andererseits aussehen.

Zunächst soll der Einfluß der funktionalen Architektur auf den kognitiven Prozeß näher betrachtet werden: Die Beschreibungsebene der funktionalen Architektur liegt zwischen der physikalischen und der repräsentationalen Beschreibungsebene. Da es für das zu entwickelnde Symbolverarbeitungsmodell keine Rolle spielen darf, wie der Prozeßablauf physikalisch realisiert ist, muß die funktionale Architektur folglich unabhängig von physikalischen Phänomenen beschrieben werden. Andererseits muß die funktionale Ebene auch unabhängig von den Inhalten der Repräsentationen sein. Es muß sich dabei also um ein nicht interpretiertes Regelschema handeln.
Trotz der genannten Einschränkungen bleibt ein gewisser Spielraum für die zu beschreibende funktionale Ebene. Man könnte die funktionale Ebene auch die *virtuelle kognitive Maschine* nennen, um es mit dem Aufbau eines Computersystems zu vergleichen. Damit wäre zu klären, wie die *relevanten primitiven Operationen* der virtuellen kognitiven Maschine aussehen ? Im nächsten Abschnitt wird Pylyshyns Kriterium für die Trennung von funktionaler Architektur und den repräsentierten intentionalen Zuständen erläutert.

4.1.4 Das Kriterium der kognitiven Beeinflußbarkeit

[9] Wie lassen sich die besagten elementaren Operationen, die die feste funktionale Architektur konstituieren, isolieren ?
Die Rede von einer festen funktionalen Architektur, wobei die einzelnen Funktionen *kognitiv unbeeinflußbar* sein müssen, heißt nicht, daß sich die elementaren Funktionen überhaupt nicht beeinflussen lassen. Vielmehr ist damit gemeint, daß sie unbeeinflußbar gegenüber einer bestimmten eingeschränkten Klasse von Faktoren sind - nämlich kognitiven Faktoren wie Änderungen in Meinungen, Zielen usw. Somit darf die feste funktionale Architektur also durch nichtkognitive Faktoren (z.B. Einnahme von Medikamenten, besondere körperliche Zustände, usw.) beeinflußt werden. Umgekehrt läßt sich Pylyshyns Bedingung der kognitiven Beeinflußbarkeit ganz allgemein so angeben:

> Eine Ein-/Ausgabefunktion ist kognitiv beeinflußbar, wenn die Veränderungen dieser Funktion durch vorhergehende Ereignisse erklärt werden können, die sich durch ein kognitives Vokabular beschreiben lassen. *Erklärt werden* heißt hier, unter einer gesetzesartigen Generalisierung subsumiert zu werden.

[9] (Pylyshyn spricht von *cognitive penetrability condition.*)

Die angegebene Bedingung für die Zuordnung einer Funktion zur funktionalen Architektur ist eine notwendige, jedoch keine hinreichende Bedingung.
Bei der empirischen Erforschung der kognitiven funktionalen Architektur sollten nach Pylyshyn die folgenden Dinge beachtet werden: Wenn Unterschiede in den kognitiven Prozessen verschiedener Menschen festgestellt werden, so sollte die angenommene innere grundlegende funktionale Architektur insofern universell sein, daß all diese kognitiven Prozesse mit ihr erklärt werden können. Dabei sollen verschiedenen Menschen lediglich unterschiedliche Inhalte in den Repräsentationen bzw. in den Algorithmen zugesprochen werden müssen. In diesem Zusammenhang lassen sich Pylyshyns Kriterien, unter denen eine bestimmte Funktion oder eine Verhaltensweise nicht der funktionalen Ebene, sondern der repräsentationalen Ebene zugeordnet werden muß, wie folgt zusammenfassen:

1. Wenn eine hypothetische Funktion von ausschließlich kognitiven Faktoren abhängt - beispielsweise unterschiedliche Instruktionen bei Versuchen oder andere Einflüsse, die zu veränderten Meinungen, Zielen usw. führen -, so gehört sie nicht der funktionalen Ebene an.

2. Wenn verschiedene Formen der hypothetischen Funktion postuliert werden müssen, um damit systematische Unterschiede in beobachteten kognitiven Phänomenen zu erklären, so gehört diese Funktion ebenfalls nicht der funktionalen Ebene an.

Die beiden obigen Bedingungen sollen also die Grenzlinie zwischen der Architektur- und der Algorithmusebene ziehen. Diese Grenzlinie ist nach Pylyshyn deshalb von so fundamentaler Bedeutung für die Kognitionswissenschaft, weil sie die Verwendung des Begriffs der *Symbolverarbeitung* als buchstäbliche Beschreibung von kognitiven Prozessen im Gegensatz zu einer bloß metaphorischen Beschreibung erst möglich macht.

In diesem Abschnitt wurde Pylyshyns methodologischer Vorschlag für die Erforschung einer *kognitiven funktionalen Architektur* dargestellt. Auf dieser kognitiven funktionalen Architektur, so Pylyshyns Ziel, läßt sich dann eine bestimmte Klasse von Phänomenen (die kognitiven Prozesse) durch formale Symbolmanipulation erklären.
Eine Einschätzung seiner Konzeption aufgrund der in dieser Arbeit angestellten Betrachtungen findet sich in Abschnitt 6.2 und in Kapitel 10.

Daß sich menschliche Intelligenzleistungen durch Symbolmanipulationen erklären lassen, ist allerdings nicht unumstritten. Insbesondere aus phänomenologischer Sicht wurden starke Einwände erhoben, welche in Kapitel 7 diskutiert werden. Es gibt aber auch noch andere Einwände. Im folgenden Abschnitt wird die Frage diskutiert, inwiefern sich die für Intelligenzleistungen relevanten Phänomene überhaupt durch symbolische Repräsentationsformalismen beschreiben lassen.

4.2 Das Problem der Wissensrepräsentation

Wie und ob überhaupt *Wissen* vollständig in formalen Systemen repräsentiert werden kann, ist durchaus umstritten. Im folgenden werden verschiedene Klassen von Wissensrepräsentationsansätzen diskutiert. Abschließend werden Argumente für die Notwendigkeit *subsymbolischer* Prozesse bzw. einer *subsymbolischen* Betrachtungsebene erörtert.

4.2.1 Verschiedene Ansätze zur Wissensrepräsentation

Traditionelle Ansätze der KI gehen von der sogenannten *Wissensrepräsentationshypothese* aus, wie sie Smith in [Smi85] formuliert. Dort schreibt Smith:

> ... any process capable of reasoning intelligently about the world must consist in part of a field of structures, of a roughly linguistic sort, which in some fashion represent whatever knowledge and beliefs the process may be said to possess ...[10]

Weiter heißt es dort:

> ... an internal process that 'runs over' or 'computes with' these representational structures ... this ingredient process is required to react only to the 'form' or 'shape' of these mental representations, without regard to what they mean or represent - this is the substance of the claim that computation involves formal symbol manipulation ...[11]

Dieser Ansatz folgt der sogenannten 'symbol system hypothesis' von Simon und Newell in [New80, NS76]. Andere Verteidiger dieses Ansatzes sind u.a. Fodor [Fod75, Fod87b], McCarthy [McC88], Minsky [Min68, Min72, Min82, Min86] und Pylyshyn [Pyl84]. Die Entwicklung des konnektionistischen Paradigmas[12] hinterfragte diesen symbolischen Ansatz zur Theorie kognitiver Prozesse. Aus konnektionistischer Sicht wird die Hypothese, daß kognitive Aktivitäten auf formale Symbolmanipulation reduziert werden können, z.B. von Dreyfus [DD87], Rumelhart et al. [RMt86] oder Smolensky [Smo88] angezweifelt. Weiterhin erweckten konnektionistische Theorien das breite Interesse an tieferen Ebenen von kognitiven und neuronalen Prozessen wieder zum Leben. Diese beiden Positionen sind nicht als vollkommen gegensätzliche in sich geschlossene Systeme zu sehen. Vielmehr zeigten Diskussionen in jüngerer Vergangenheit eine differenziertere Sichtweise. In Anlehnung an Rumelhart et al. [RMt86], Fodor und Pylyshyn [FP88] und Smolensky [Smo88] lassen sich die folgenden Sichtweisen zum Wissensrepräsentationsproblem unterscheiden:

1. zusammengesetzt-symbolische Ansätze
2. lokale nicht-zusammengesetzt-symbolische Ansätze

[10] In Brachman & Levesque [BL85] Seite 33.
[11] [ibid.]
[12] Vergleiche Abschnitt 2.8 und 8.4.

3. verteilte nicht-zusammengesetzt-symbolische Ansätze
4. kognitiv-subsymbolische Ansätze
5. neuronal-subsymbolische Ansätze

Der erste Punkt entspricht Smiths *knowledge representation hypothesis*. Nach Fodor und Pylyshyn[13] kann man sagen, daß diese Ansätze durch die Zusammensetzbarkeit der Bedeutung charakterisiert werden können. Durch syntaktische Merkmale des Repräsentationssystems lassen sich dabei atomare von zusammengesetzten Ausdrücken unterscheiden. Die zusammengesetzten Ausdrücke lassen sich aus atomaren Ausdrücken nach bestimmten syntaktischen Regeln rekursiv aufbauen. Dabei ist die Bedeutung der zusammengesetzten Ausdrücke eine Funktion seiner syntaktischen Struktur und der Bedeutung der atomaren Ausdrücke in ihnen. Hierbei läßt sich annehmen, wie beispielsweise Frixione et al. [FSG89], daß sich zu *jedem* derartigen System eine modelltheoretische Semantik der syntaktischen Ausdrücke angeben läßt, etwa nach dem Vorbild der Tarskischen Semantik für die Prädikatenlogik [Tar36] oder der Kripke Modelle für intensionale Logiken [Kri59].

Die Ansätze, die unter den zweiten Punkt fallen, sind insofern symbolisch zu nennen, als daß bei ihnen jedes Element des Repräsentationssystems ein Symbol ist. Jedem Element ist eine Bedeutung zugeordnet. Diese Bedeutung ist an ein einzelnes Element gebunden und nicht über mehrere Elemente - wie im dritten der obigen Fälle - verteilt. Bei den Ansätzen des zweiten Typs kann das Zusammensetzungsprinzip aus dem folgenden Grund nicht angewendet werden: Es gibt keine syntaktischen Regeln, die komplexe Ausdrücke erzeugen können; vielmehr ist jedes Symbol in gewissem Sinn atomar. Damit kann man für diese Ansätze auch keine modelltheoretische Semantik angeben. Es sei denn in einem sehr banalen Sinn, wobei jedem Symbol ein eigenes Referenzobjekt zugeordnet wird. *Lokale konnektionistische Modelle* bei denen jede begriffliche Einheit durch einen eigenen Knoten des Netzwerks repräsentiert wird, sind von diesem Typ. Gerade in der fehlenden Zusammensetzbarkeit der Bedeutung sehen Fodor und Pylyshyn den charakteristischen Unterschied zwischen konnektionistischen und klassischen Ansätzen.[14]
In dem verteilten, dritten der aufgeführten Fälle, ist jede begriffliche Einheit durch ein Aktivierungsmuster mehrerer Netzwerkknoten repräsentiert. Andererseits ist auch jeder Netzwerkknoten an der Repräsentation mehrerer begrifflicher Einheiten beteiligt. Hier läßt sich jedem Knoten des Netzwerks ein eigenes 'Mikromerkmal' zuordnen, aus denen sich die obengenannten begrifflichen Einheiten zusammensetzen. Solche Modelle sind zwar in gewissem Sinn *subbegrifflich*, jedoch nicht subsymbolisch. Begriffliche Einheiten sind über mehrere Netzwerkknoten, mithin über mehrere physikalische Einheiten verteilt, jedoch läßt sich jedem Netzwerkknoten eine eigene Bedeutung zuschreiben.[15]

[13] In Fodor & Pylyshyn [FP88] weisen sie darauf hin, daß wir kognitive Fähigkeiten haben, die gerade eine rekursive Zusammensetzbarkeit von Strukturen wiederspiegeln; z.B. daß wir im Prinzip beliebig lange Sätze grammatikalisch korrekt erzeugen oder verstehen können.

[14] Siehe hierzu Fodor & Pylyshyn [FP88], Seite 15-19.

[15] Siehe beispielsweise Lange [LD89]. Diederich [Die91] bietet einen Überblick.

Bei Modellen des vierten Punktes werden die subsymbolischen Einheiten nicht als bedeutungstragende Elemente angesehen, wie im dritten Falle die Mikromerkmale. Man spricht ihnen auch keine genaue physiologische Korrespondenz zu. Hier müssen die subsymbolischen Einheiten, also die Knoten eines Netzwerks, eher als bloße theoretische Konstrukte angesehen werden, die dazu dienen das Verhalten auf der kognitiven Beschreibungsebene zu bestimmen. Das heißt, mit formalen Mitteln - also mittels Computerprogrammen - zu berechnen.[16] Fodor und Pylyshyn haben diesen vierten Punkt nicht betrachtet. Ihnen scheinen nur der dritte und fünfte Punkt als mögliche Alternativen in Frage zu kommen. Unter dem fünften Punkt lassen sich die neuronalen konnektionistischen Modelle zusammenfassen. Bei solchen Netzwerken sind Repräsentationen über nicht-symbolische Einheiten verteilt, von denen man annimmt, daß ihnen genau korrespondierende anatomische und neurophysiologische Prozesse zugeordnet werden können. Fodor und Pylyshyn [FP88] sehen solche Ansätze als unwichtig für die Betrachtung bewußter kognitiver Phänomene an, da der Fokus dieser Ansätze nur die darunterliegenden Prozesse betrifft. Darüberhinaus läßt sich über solche Modelle zwar theoretisieren, jedoch scheint die Verwirklichung solcher Modelle als sehr schwierig.
Zusammenfassend kann man sagen, daß die Modelle unter den Punkten 1. bis 3. als symbolische Ansätze gelten können, während die Modelle unter den Punkten 4. und 5. als subsymbolische Ansätze angesehen werden können. Hingegen lassen sich die Modelle unter 3. bis 5. als verteilte Modelle im Gegensatz zu den Modellen unter 1. und 2. einordnen. Letztlich können die Modelle unter 1. bis 4. als Ansätze auf der kognitiven Ebene eingestuft werden, während die Modelle unter 5. als subkognitive (physiologische) Modelle gelten können.

4.2.2 Ist eine subsymbolische Ebene notwendig ?

Die Bedeutung zusammengesetzt-symbolischer Modelle kognitiver Prozesse sollte nicht unterschätzt werden, selbst wenn ihr Wert nur daher rührt, daß nicht zusammengesetzte Ansätze bisher keine plausible Alternative lieferten.[17] Trotzdem gibt es beim zusammengesetzten symbolischen Ansatz theoretische Überlegungen, die zur Postulierung einer subsymbolischen Ebene führen. Auf der einen Seite scheint einigen Autoren das Problem der Referenz - auch im Fall der zusammengesetzten symbolischen Ansätze - eine Ebene subsymbolische Prozesse erforderlich zu machen.[18] Fodor [Fod80] geht zwar in einer solipsistischen Sichtweise davon aus, daß die einzigen funktionalen Beziehungen, die für eine präzise Theorie des Geistes erforderlich sind, lediglich diejenigen sind, die zwischen den Symbolen innerhalb eines Modells kognitiver Prozesse bestehen. Somit sind alle semantischen Aspekte, wie der der Referenz für eine solche Theorie irrelevant. Fodors These ist allerdings nicht unumstritten. So schreibt beispielsweise Harman:

[16] Hingegen vertreten Smolensky [Smo88] und Hofstadter [Hof85] diese Position.
[17] Siehe z.B. Fodor & McLaughlin [FM90].
[18] Siehe beispielsweise Frixione et al. [FSG89].

> of primary importance are functional relations to the external world in connection with perception, on the one hand, and action, on the other[19]

Harman geht dabei allerdings nicht näher auf diese Beziehungen ein. Auch Sloman und Cohen [SC86] sehen eine Notwendigkeit, das Phänomen der Referenz von Symbolen zu klären. Frixione et al. [FSG89] versuchen diese Verbindung der Symbole zu dem auf das sie referieren, auf einer subsymbolischen Ebene zu erklären. Es wurde oben angedeutet, wie eine modelltheoretische Semantik für Modelle des ersten Typs angegeben werden kann, wobei die Modelltheorie allerdings nicht für sich in Anspruch nimmt, eine empirisch adäquate Theorie für die Referenz in kognitiven Prozessen sein zu können. Frixione et al. sehen darüber hinaus auch einige a priorische Grenzen der Modelltheorie. Die empirische Inadäquatheit zunächst hängt mit der Tatsache zusammen, daß Menschen die Referenz komplexer symbolischer Gebilde nicht nach den Regeln ihres syntaktischen Aufbaus zu bilden scheinen. Dies wird allerdings bereits spätestens durch Wittgensteins Spätphilosophie, seinen Philosophischen Untersuchungen [Wit53] deutlich. Das Problem klingt aber auch schon im Tractatus an. In letzterem heißt es:

> Der Mensch besitzt die Fähigkeit Sprachen zu bauen, womit sich jeder Sinn ausdrücken läßt, ohne eine Ahnung davon zu haben, wie und was jedes Wort bedeutet. - Wie man auch spricht, ohne zu wissen, wie die einzelnen Laute hervorgebracht werden.
>
> Die Umgangssprache ist ein Teil des menschlichen Organismus und nicht weniger kompliziert als dieser.
>
> Es ist menschenunmöglich, die Sprachlogik aus ihr unmittelbar zu entnehmen. Die Sprache verkleidet den Gedanken. Und zwar so, daß man nach der äußeren Form des Kleides nicht auf die Form des bekleideten Gedankens schließen kann; weil die äußere Form des Kleides nach ganz anderen Zwecken gebildet ist als danach, die Form des Körpers erkennen zu lassen.[20]

Mit den a priori Grenzen der modelltheoretischen Semantik sind allerdings Grenzen gemeint, die gleichermaßen für nicht antropomorphe, intelligente Wesen gelten würden. Frixione et al. [FSG89] weisen hierbei auf das folgende Problem hin: Wenn eine Interpretationsfunktion angebbar sein soll, die Werte für komplexe Ausdrücke bestimmt, so muß sie in gewissem Sinn 'gewußt' werden können. Da es sich dabei um Funktionen handelt, die Symbole auf Objekte der Welt abbilden, kann eine solche Funktion nicht ausschließlich durch Symbole beschrieben werden, da dies einen unendlichen Regreß implizieren würde. Daher, so schließen Frixione et al. [FSG89], muß es für die Interpretation zumindest von einigen Symbolen des Systems (den atomaren Ausdrücken) eine 'subsymbolische' Realisierung geben. Eine derartige Realisierung könnte nach Frixione et al. durch Ein-/Ausgabenetzwerkknoten in einem konnektionistischen Modell geschehen, denen keine konkrete Bedeutung zugeordnet werden kann.[21]

[19] In Harman [Har87] Seite 67.

[20] In Wittgenstein [Wit21] Absatz 4.002.

[21] Die Idee, die Bedeutung von systemintern verwendeten Symbolen durch die Rezeptoren und Effektoren eines Systems in seiner Umwelt dem System selbst zu überlassen, hat eine ganze Reihe von Anhängern

Andere Gründe als Probleme der Referenz sieht Smolensky [Smo88, Smo90] für die Postulierung einer subsymbolischen Ebene:

- Existierende KI-Systeme die auf der *symbol system hypothesis* beruhen, scheinen zu inflexibel zu sein, um wirkliches menschliches Wissen und Können modellieren zu können.
- Die Formulierung von Expertenwissen durch Regeln scheint für viele wichtige Bereiche unpraktikabel zu sein (z.B. für das Alltagswissen).
- Durch die *symbol system hypothesis* hat man bisher so gut wie überhaupt keinen Aufschluß darüber erhalten, wie Wissen im menschlichen Gehirn repräsentiert ist.

Die eigentliche Motivation Smolenskys für die Suche nach konnektionistischen Alternativen ist die Hoffnung, daß man durch solche Alternativen den Zielen der Kognitionswissenschaft näher kommt. Hierbei meint Smolensky insbesondere, daß man Modelle entwickelt, die *intuitive* kognitive Aktivitäten erklären. Dazu zählt er alle Aktivitäten außer die des bewußten Regelfolgen; zum Beispiel Sprechen in der Muttersprache, das Urteilen in alltäglichen Situationen oder kreative Tätigkeiten, wie Dichten oder auch das Beweisen eines mathematischen Satzes. Ein bewußtes Regelfolgen könnte das Kochen nach den Rezepten eines Kochbuches sein oder die Anwendung eines bisher noch nicht geläufigen Rechenverfahrens. Smolensky sieht das Wissen, das in konnektionistischen Modellen enthalten ist, in den unterschiedlich starken Verbindungen zwischen den einzelnen Knoten eines Netzwerkes. Somit schlägt er die folgende Hypothese als Alternative zur *symbol system hypothesis* zur Erklärung, Beschreibung und Modellierung intuitiver kognitiver Aktivitäten vor:

(8) a. The connectionist dynamical system hypothesis:

> The state of the intuitive processor at any moment is precisely defined by a vector of numerical values (one for each unit). The dynamics of the intuitive processor are governed by a differential equation. The numerical parameters in this equation constitute the processor's program or knowledge. In learning systems, these parameters change according to another differential equation.[22]

Dazu schlägt er die folgende korrespondierende Hypothese vor, die den Zusammenhang zwischen den unbewußten Prozessen und den bewußt werdenden kognitiven Phänomenen zumindest im Prinzip erklären können soll:

(8) b. The subconceptual unit hypothesis:

gefunden. Vergleiche hierzu Hanard [Han87] oder Wrobel [Wro91]. In der philosophischen Literatur finden sich hierzu beispielsweise Arbeiten von Johnson [Joh87] oder von Lakoff [Lak87]. Vergleiche auch Abschnitt 8.5 über Selbstorganisation.

[22] In Smolensky [Smo88] Seite 6.

> The entities in the intuitive processor with the semantics of conscious concepts of the task domain are complex patterns of activity over many units. Each unit participates in many such patterns.[23]

Smolensky sieht zumindest als ein wesentliches Ziel der Kognitionswissenschaft die Entdeckung der Prinzipien an, die den folgenden Dingen zugrunde liegen:

1. Die 'subbegriffliche' Repräsentation in unterschiedlichen Problembereichen.
2. Die Vorgänge, die die Gewichtungen der verschiedenen Verbindungen in konnektionistischen Netzwerken verändern und dabei das Gesamtverhalten des Netzwerkes an äußere Anforderungen anpaßt; also die Lernprozesse in Netzwerken.
3. Der angemessenen Wahl von Beschreibungsmerkmalen der Phänomene in dem jeweils zu modellierenden Problembereich.

Smolensky geht davon aus, daß die Lernprozesse in konnektionistischen Netzwerken selbsttätig die einzelnen 'subbegrifflichen' Repräsentationen erzeugen, die für die äußeren Anforderungen angemessen sind. Somit würde man vermutlich aus den Erkenntnissen zum zweiten Punkt wichtige Rückschlüsse auf den ersten Punkt machen können.
Weiterhin erhofft sich Smolensky aus den noch zu findenden allgemeinen Prinzipien der 'subbegrifflichen' Betrachtungs- und Beschreibungsebene, Aufschlüsse über die neuronalen Prozesse, die kognitive Phänomene hervorbringen. Die Paralellen zwischen der 'subbegrifflichen Ebene' und der neuronalen Ebene sieht er darin, daß in beiden Fällen Berechnungen an sehr vielen Stellen gleichzeitig und insgesamt hochkompliziert und dynamisch ablaufen. Smolensky will hiermit zunächst einmal Forschungsziele für die Kognitionswissenschaft unter einem konnektionistischen Paradigma aufstellen. Smolenskys Ziel ist es somit, die *allgemeinen Prinzipien* zu entdecken, die unter anderem dem, was Pylyshyn als die 'kognitive funktionale Architektur' bezeichnet, zugrunde liegen. Denn die spezifische Art und Weise der Symbolmanipulation, bei der die Symbole intentionale Einstellungen repräsentieren, muß man wohl auch zu den intuitiven kognitiven Fähigkeiten zählen. Pylyshyn ordnet die durch Symbole repräsentierten intentionalen Einstellungen auf der intentionalen Ebene an. Die Art und Weise, wie diese Symbole manipuliert werden, wird von der *kognitiven funktionalen Architektur* bestimmt.

[23] ibid.

Teil II

Methodologische Untersuchungen

Der zweite Teil der Arbeit befaßt sich mit methodologischen Problemen der künstlichen Intelligenz bzw. der Kognitionswissenschaft.
Die beiden Wissenschaften haben einen untereinander eng verwandten Phänomenbereich zum Gegenstand. Insbesondere scheint in den angesprochenen Phänomenbereichen eine erhebliche Komplexität vorzuherrschen.
In Kapitel 6 wird dies mit Hilfe des Begriffs der Kolmogoroffkomplexität ausdrücklich thematisiert. Dabei wird die Diskrepanz zwischen der Einfachheit der (algorithmischen) Beschreibung von behaupteten Prinzipien bestimmter kognitiver Prozesse einerseits und der Komplexität von zu erklärenden oder zu beschreibenden Phänomenen andererseits betont. Aus dieser Diskrepanz erwächst einerseits die Frage, wieviel mit der Erkenntnis derartiger 'Prinzipien' gewonnen wird. Andererseits stellt sich aber auch die Frage, was Wissenschaftler überhaupt dazu bewegen kann, etwas Bestimmtes als die Prinzipien auszuzeichnen. Die Erörterung dieser Frage in Kapitel 6 führt auf universalistische Vorstellungen des/der Wissenschaftler(s) über den jeweiligen Phänomenbereich der KI bzw. der Kognitionswissenschaft, der zunächst immer nur durch einzelne Beispiele gegeben ist. Daher bietet Kapitel 5 im ersten Abschnitt einen historischen Überblick über das philosophische Universalienproblem. In den weiteren Abschnitten von Kapitel 5 wird der Bezug der Universalientheorien zu möglichen Prinzipien der KI respektive von Kognitionen erörtert.

Kapitel 5

Universalien und Prinzipien von Intelligenz

Intelligenz, Intelligenzleistungen, kognitive Prozesse sind abstrakte, allgemeine Begriffe. Will man den Gegenstandsbereich der künstlichen Intelligenz oder der Kognitionswissenschaft abgrenzen, so müssen die Subsumptionsbereiche der genannten Begriffe geklärt werden.
Will man *Prinzipien* von Intelligenz, künstlicher Intelligenz oder von kognitiven Prozessen finden, so ist die Klärung der obigen Begriffe ebenfalls von entscheidender Bedeutung, wie in Abschnitt 5.3 dargelegt wird.
Aus diesem Grund skizziert Abschnitt 5.1 zunächst die Geschichte der philosophischen Universaliendiskussion. Abschnitt 5.2 stellt eine abstrakte Fassung des Intelligenzbegriffs oder - genauer - von Intelligenzleistungen vor, auf der Abschnitt 5.3 aufbaut. Abschnitt 5.4 letztlich diskutiert den Status des Begriffs von Intelligenzleistungen vor den verschiedenen Universalientheorien und dessen Konsequenzen für den möglichen Erfolg bei der Suche nach den Prinzipien von Intelligenz.

5.1 Geschichte des Universalienproblems

Problemgeschichtlich geht die Diskussion um Universalien auf Aristoteles zurück, der Platons Universalienkonzeption kritisierte. Während der mittelalterliche Universalienstreit primär theologisch motiviert war, zeigte sich im britischen Empirismus eine durch die naturwissenschaftlichen Fortschritte stimulierte Diskussion.
Die Gruppen von Universalientheorien des *Realismus, Konzeptualismus, Nominalismus* und der*Ähnlichkeitstheorien* werden im folgenden skizziert.

5.1.1 Realismus

Der Realismus hält Universalien für real existent. Die beiden Hauptfassungen des Realismus gehen auf Platon und Aristoteles zurück. Platons Theorie war die frühere der beiden.

Die Probleme seiner Theorie, erkannte Platon zum Teil selbst, zum anderen Teil wurden sie von Aristoteles aufgeworfen.

Platon

Platons Interesse an der Universalienproblematik wurde durch seinen Lehrer Sokrates geweckt. Sokrates interessierte sich hauptsächlich für die menschlichen Tugenden, wobei er nach einer sicheren und genauen Definition dieser Tugenden suchte. Sokrates meinte, daß beispielsweise Gesundheit, Größe oder Stärke in all ihren Instanzen jeweils das Gleiche sein müsse. Platon weitete diese Sichtweise nicht nur auf Gegenstände, sondern auch auf Adjektive wie *rot* oder *schön* aus. Es müsse auch etwas wie *Röte* existieren, das allen roten Dingen gemeinsam ist.
Platon entwickelte daraus seine *Theorie der Formen*, welche jeder Universalie eine eigene singuläre Substanz zuschrieb. Diese Substanz sollte zeitlos und unabhängig von ihren konkreten Manifestationen in Gegenständen existieren. Im Gegensatz zu singulären Gegenständen können die Formen nur durch den Intellekt erfaßt werden. Die Seele erfährt vor der Geburt eine *Ideenschau*, bei der sie alle Formen kennenlernt. Später erinnert sie sich an die Formen aufgrund der sinnlichen Wahrnehmung von Einzeldingen.
Bei seiner Theorie der Formen stieß Platon auf das folgende Problem:[1] Wenn eine Form zu ihren Einzeldingen ein solches Verhältnis hat, daß die Einzeldinge unvollkommene Kopien der Form sind, so entsteht dabei ein unendlicher Regress. Wenn die Form F sowohl auf sich selbst als auch auf ihre Einzeldinge prädizierbar ist, so muß F und ihren Einzeldingen etwas gemeinsam sein. Zum Beispiel ist die Form *Röte* auf alle roten Dinge prädizierbar, als auch auf sich selbst - schließlich ist *Röte* ja das Paradeobjekt der roten Objekte und zweifelsohne ebenfalls *rot*. Also muß es eine Form F' geben, die genau dieses Gemeinsame ist. F' ist dann ihrerseits auf sich selbst prädizierbar, so daß es eine weitere Form F'' geben muß, die ... etc.
Platon gelang es nicht, eine schlüssige Erklärung für das Verhältnis zu geben, in dem die Formen zu ihren Einzeldingen stehen. Dieses Problem scheint generell unlösbar zu sein.

Aristoteles

Aristoteles kritisierte Platons Theorie der Formen und stellte dem seine eigene Universalientheorie gegenüber, bei der er die von Platon behauptete Präexistenz der Universalien ablehnte.
Die wichtigsten Argumente, die Aristoteles gegen Platons Theorie der Formen hervorbrachte, waren die drei folgenden:

- Platons Formen als separate vollkommene Substanzen stellen eine unnötige Verdopplung der Substanzen dar. Dadurch wird nichts erklärt, das eigentliche Problem wird lediglich verschoben.

[1] Platons Parmenides [Pla] 132 ff.

- Bei Platon werden die Kategorien der Substanz und der Eigenschaften nicht strikt voneinander getrennt. Formen sind gleichzeitig als Eigenschaften ihrer Einzeldinge und als individuelle Substanzen gedacht. Nach Aristoteles aber sind Substanzen Einzeldinge und *haben* Eigenschaften - sie können damit aber keine Eigenschaften *sein*.

- Das sogenannte 'dritter Mensch'[2] Argument, bei dem Aristoteles argumentiert, daß ein einzelner Mensch und die Idee des *Menschen* doch etwas Gemeinsames haben müsse, es müsse noch einen 'dritten Menschen' geben, um die Beziehung zwischen ihnen zu klären. Dies ist der bereits genannte unendliche Regress von Ideen, den auch Platon selbst entdeckte.

Damit sind bei Aristoteles Universalien keine Substanzen, die unabhängig von den Einzeldingen existieren. Die Universalien existieren lediglich als gemeinsame Elemente in den Einzeldingen. Die Einzeldinge können gemäß gemeinsamer Elemente, die sie beinhalten, in Klassen eingeteilt werden. Insofern ist Aristoteles Theorie ökonomischer als Platons, da sie für die Klassifizierung sinnlich wahrnehmbarer Objekte nur *eine* Welt von Entitäten erfordert. Der Unterschied zwischen den beiden Theorien zeigte sich auch in den Namen, die sie in der Scholastik hatten: Platons Theorie war eine Theorie der *universalia ante rem* (Universalien unabhängig von den Einzeldingen), während Aristoteles Ansatz eine Theorie der *universalia in rebus* (Universalien in den Dingen) war.
Abgesehen von idealen Begriffen, wie geometrische Begriffe, die nach Platon keine wirklichen Instanzen haben, scheint Aristoteles Theorie die menschliche Erfahrung besser zu erklären. Schließlich scheint intuitiv ein Einzelding wirklich eine Instanz seiner Universalien zu sein. So sagen wir im allgemeinen auch, daß die Tomate oder der Feuerwehrwagen rot ist und nicht, daß die Tomate erfolglos versucht rot zu sein (der Röte nahe zu kommen), wie es Platons Theorie sagen würde.
Aristoteles sah die Aneignung der Universalien als einen allmählichen Prozeß an, der durch sinnliche Wahrnehmung und der Erinnerung an solche stimuliert wird. Somit ist der Zugang zu den Universalien auch für Aristoteles durch den Intellekt gegeben, jedoch geschieht dies dadurch, daß der Intellekt nach und nach mit Zunahme der sinnlichen Wahrnehmung die Begriffe schärft. So lernen wir beispielsweise den Begriff von der Zahl 2 dadurch, daß wir mit Paaren von Dingen konfrontiert werden. Daß $2 + 2 = 4$ gilt, lernen wir dadurch, daß beispielsweise zwei Paare von Äpfeln vier Äpfel sind. Mit der Zeit erkennen wir, daß die Zahl 2 für ein beliebiges Paar von Dingen steht. Wir erkennen, daß $2 + 2 = 4$ eine notwendige Wahrheit ist, die auf beliebige zwei Paare von Dingen anwendbar ist.

Kritik am Realismus

Aristoteles' Version des Realismus zeigte sich als widerstandsfähiger gegen Kritik als Platons Theorie. Nichtsdestotrotz wurden im Laufe der Zeit, insbesondere in der britischen Tradition des Empirismus auch an Aristoteles' Realismus schwere Einwände laut.

[2]Siehe Aristoteles: Metaphysik [Ari] (A), 990.17.

5.1.2 Konzeptualismus

Der Konzeptualismus ist im wesentlichen dem britischen Empirismus zuzuschreiben, welcher von einer sensualistischen Erkenntnistheorie ausgeht.

> *Nichts ist im Verstand, was nicht vorher in den Sinnen war.*

Die Lösung des Universalienproblems sucht der Konzeptualismus im erkennenden Subjekt, statt in einer äußeren, 'realen' Welt.
Der Konzeptualist erklärt die Allgemeinheit eines Wortes dadurch, daß er auf allgemeine Begriffe verweist, zu denen die Wörter korrespondieren. Dabei muß geklärt werden, was ein solcher allgemeiner Begriff ist.
Für Locke ebenso wie für Hobbes gehört das Universale nicht zur realen Existenz der Dinge. Vielmehr handelt es sich dabei um Produkte des Verstands. Locke betonte, daß nahezu alles Denken verbal ist; nonverbale Vorstellungen beim Denken treten nur in einem sehr beschränkten Ausmaß auf. So deuten bestimmte Stellen in seinem *Essay on Human Understanding* [Loc90] daraufhin, daß er Ideen nicht nur als Bilder ansah, die zu bestimmten Worten korrespondieren. Er sah sie auch als Bedeutung von Wörtern an. Eine Idee von einem Wort, z.B. von *Tisch* zu haben, heißt das Wort *Tisch* zu verstehen und in bestimmter Weise zu gebrauchen. Eine korrekte Idee eines Wortes zu haben, heißt das Wort auf gleiche Weise zu gebrauchen, wie es die anderen tun.
Berkeley wandte sich gegen die Annahme '*unum nomen unum nominatum*', daß immer das gleiche Wort auch mit der gleichen Vorstellung verknüpft sei. Berkeley reduziert das Allgemeine auf Regeln des Gebrauchs genereller Ausdrücke.[3]
Er zeigte, daß Worte wie *Kraft* trotzdem sinnvoll benutzt werden können, selbst wenn Fragen, wie '*Was ist Kraft ?*' (in der Physik) zu keiner Antwort führen.[4] Damit durchbrach Berkeley die Mauern des strikten Empirismus und antizipierte den Theorieaufbau der modernen Wissenschaft, insbesondere der modernen Physik.
Wie wir zu universellen Vorstellungen kommen, erklärte Hume durch die Gewöhnung des Menschen an die Assoziation bestimmter Ideen mit bestimmten Worten. Man könnte nach Hume sagen, daß wir denken lernen, indem wir sprechen lernen und nicht umgekehrt. Und das Erlernen der Sprache geschieht hauptsächlich durch Gewohnheit.[5]

Eine Konsequenz eines solchen Konzeptualismus ist, daß Begriffe - im Gegensatz zu einer einfachen realistischen Theorie - Veränderungen unterworfen sein können. Begriffe unterliegen einer Entwicklung und Änderung, wenn sich herausstellen sollte, daß es nützlich ist, Klassifikationen anders als bisher zu treffen.

[3] Hier nähert sich Berkeley den Auffassungen des späten Wittgenstein 'Die Bedeutung eines Wortes ist sein Gebrauch in der Sprache.'
[4] In Berkeley [Ber10].
[5] Siehe dazu Hume [Hum48].

5.1.3 Nominalismus

Der Nominalismus behauptet gegenüber dem Konzeptualismus sogar, daß nur Wörter - nicht Begriffe - allgemein sind. Sowohl nominalistische als auch konzeptualistische Theorien versuchen zu erklären, wie *Wörtern* Allgemeinheit zukommen kann und wie sie überhaupt Bedeutung haben können. Der Nominalismus wurde bereits im scholastischen Universalienstreit unter Anderen von Peter Abälard und William von Ockham vertreten. Im siebzehnten Jahrhundert wurde er erneut von T. Hobbes in seinem *Leviathan* [Hob51] diskutiert; für ihn waren Gattungs- und Artbegriffe 'allgemeine Benennungen'. Außer Benennungen gibt es nichts Allgemeines. Nach Ockham sind Universalien Ausdrücke, die entweder auf einzelne Objekte oder auf Mengen von Objekten referieren. Jedoch kann man ihnen keine eigene Existenz zusprechen. Universalien sind vielmehr Prädikate oder Bedeutungen, die lediglich einen *logischen* Status haben und die für Denken und Sprechen notwendig sind.
In seiner extremen Form, erscheint der Nominalismus so unhaltbar, daß vermutlich niemand eine solche Position jemals vertrat. Er würde behaupten, daß eine Klasse von Einzeldingen, beispielsweise Tische, nichts gemeinsam haben, außer die Tatsache, daß sie alle *Tische* genannt werden.
Der Nominalismus muß sich auf eine *Ähnlichkeitstheorie* reduzieren. Denn der Nominalismus akzeptiert nur die Existenz von Einzeldingen und läßt die individuellen Eigenschaften der Einzeldinge auch nur ihnen zukommen. Daher benötigt der Nominalismus für die Universalität von Worten Ähnlichkeitsbeziehungen zwischen den Einzeldingen.

5.1.4 Ähnlichkeitstheorien

Locke vertrat eine Ähnlichkeitstheorie, die einem aristotelischen Realismus sehr nahe kommt. Der Ähnlichkeitsgrad zweier Objekte hängt bei ihm von dem Maß der *qualitativen* Übereinstimmung ab, die mehr oder weniger direkt dem aristotelischen Charakeristika entspricht. Bei einer reinen Ähnlichkeitstheorie können zwei Objekte einander ähneln. Dabei muß es zwar einen bestimmten Aspekt geben, unter dem sie einander ähnlich sind, jedoch muß dieser Aspekt nicht als das identische Etwas betrachtet werden, das beiden Objekten zukommt. Somit läßt sich ein aristotelischer Realismus vermeiden. Gegen Ähnlichkeitstheorien wurde oft das Argument vorgebracht, daß Ähnlichkeit seinerseits eine Universalie sein muß.
B. Russell argumentierte in *Problems of Philosophy* (1912) [Rus12]: Wenn wir Universalien wie *weiß* oder *dreieckig* vermeiden wollen, so müssen wir auf bestimmte *weiße* bzw. *dreieckige* Objekte verweisen. Wir können dann sagen, daß etwas *weiß* bzw. *dreieckig* ist, wenn es den ausgewählten Objekten im richtigen Maße ähnelt. Jedoch muß dann diese Ähnlichkeit selbst eine Universalie sein, anderenfalls gerät man in einen unendlichen Regreß. Man müßte Ähnlichkeiten zwischen Ähnlichkeitsbeziehungen behaupten, aufgrund derer bestimmte Objektpaare einander ähneln, während andere einander nicht ähneln.
R. Carnap entwickelte in *Der logische Aufbau der Welt* (1928) [Car28] eine Theoriesprache, in der als einziger Grundprädikator die zweistellige *Ähnlichkeitserinnerung* zwischen

Elementarerlebnissen vorgesehen ist. Price [Pri53] weist darauf hin, daß der von Russell angezeigte unendliche Regreß ruhig bestehen könne - er müsse gar nicht (durch die Postulierung von Universalien) aufgelöst werden. Der Wert der Ähnlichkeitstheorie bestehe darin, daß sie nicht wie die Realismustheorien Erklärung und Beschreibung durcheinander bringt. Über uns selbst ist die Frage nach den Universalien eine Frage der *Erklärung*. Über die Welt ist die Frage nach Universalien eine Frage der *Beschreibung*. Auch wurden Zweifel an dem Sinn der Annahme laut, daß die Ähnlichkeit zwischen Objekten immer schon besteht, und später nur entdeckt werden muß.[6]

Wittgenstein führte im zwanzigsten Jahrhundert den Begriff der *Familienähnlichkeit* ein. Er meinte damit ein ganzes Netz von Ähnlichkeiten, die zwischen Objekten gleichen Namens bestehen. Dabei bestehen Ähnlichkeiten in unterschiedlicher Hinsicht kreuz und quer zwischen den einzelnen Objekten der fraglichen Klasse. Er nennt selbst als Beispiel die Klasse der Spiele; Brettspiele, Kartenspiele, Ballspiele, ... In den Philosophischen Untersuchungen heißt es:

> 66. Betrachte z.B. einmal die Vorgänge, die wir $\succ$ Spiele $\prec$ nennen. Ich meine Brettspiele, Kartenspiele, Ballspiele, Kampfspiele, usw. Was ist allen gemeinsam ? - Sag nicht: "Es *muß* ihnen etwas gemeinsam sein, sonst hießen sie nicht $\succ$ Spiele $\prec$" - sondern *schau*, ob ihnen allen etwas gemeinsam ist. - Denn wenn Du sie anschaust, wirst Du zwar nicht etwas sehen, was *allen* gemeinsam wäre, aber du wirst Ähnlichkeiten, Verwandtschaften, sehen, und zwar eine ganze Reihe. Wie gesagt: denk nicht, sondern schau ! - Schau z.B.- die Brettspiele an, mit ihren mannigfachen Verwandtschaften. Nun geh zu den Kartenspielen über: hier findest du viele Entsprechungen mit jener ersten Klasse, aber viele gemeinsame Züge verschwinden, andere treten auf. Wenn wir nun zu den Ballspielen übergehen, so bleibt manches Gemeinsame erhalten, aber vieles geht verloren. - Sind sie alle *unterhaltend* ? Vergleiche Schach mit dem Mühlfahren. Oder gibt es überall ein Gewinnen und Verlieren, oder eine Konkurrenz der Spielenden ? Denk an die Patiencen. In den Ballspielen gibt es Gewinnen und Verlieren; aber wenn ein Kind den Ball an die Wand wirft und wieder auffängt, so ist dieser Zug verschwunden. Schau, welche Rolle Geschick und Glück spielen. Und wie verschieden ist Geschick im Schachspiel und Geschick im Tennisspiel. Denk nun an die Reigenspiele: Hier ist das Element der Unterhaltung, aber wie viele der anderen Charakterzüge sind verschwunden ! Und so können wir durch die vielen, vielen anderen Gruppen von Spielen gehen. Ähnlichkeiten auftauchen und verschwinden sehen.
>
> Und das Ergebnis dieser Betrachtung lautet nun: Wir sehen ein kompliziertes Netz von Ähnlichkeiten, die einander übergreifen und kreuzen. Ähnlichkeiten

[6] Neben dem späten Wittgenstein und Goodman (siehe [Goo69]) zweifelt beispielsweise auch M. Black daran, wie das folgende Zitat aus [Bla62] Seite 37 zeigt: *It would be more illuminating in some of these cases to say that a metaphor creates the similarity than to say it formulates some similarity antecedently existing.*

> im Großen und Kleinen.
>
> 67. Ich kann diese Ähnlichkeiten nicht besser charakterisieren als durch das Wort 'Familienähnlichkeiten'; denn so Übergreifen und kreuzen sich die verschiedenen Ähnlichkeiten, die zwischen den Gliedern einer Familie bestehen: Wuchs, Gesichtszüge, Augenfarbe, Gang, Temperament, etc. etc. - Und ich werde sagen: Die $\succ$ Spiele $\prec$ bilden eine Familie. ...[7]

Im zwanzigsten Jahrhundert erlebte der Universalienstreit eine überraschende Neuauflage, (moderner Universalienstreit) der durch die Grundlagenkrise der Mathematik ausgelöst wurde. Eine platonische Position würde den Umgang mit Unendlichkeiten erleichtern, und wurde unter anderem von G. Cantor, G. Frege, dem frühen B. Russell, K. Gödel und A. Church vertreten.
Doch birgt er auch Probleme wie es beispielsweise in der Russellschen Mengenantinomie zum Ausdruck kommt.[8] Dadurch und durch weitere Probleme, wurde die Entwicklung von nominalistischen Systemen stimuliert. So haben beispielsweise Goodman & Quine [GQ47] ein nominalistisches System für einen Bereich der Mathematik entwickelt.

5.2 Zur Entdeckung der Prinzipien von Intelligenz

Als *Prinzip* wird gemeinhin dasjenige bezeichnet, wovon etwas in irgendeiner Weise seinen Ausgang nimmt, sei es dem Sein oder dem Geschehen oder der Erkenntnis nach. Bei dem methodischen Aufbau von Einzelwissenschaften sind die Prinzipien grundlegende Einsichten bzw. Aussagen, auf denen systematisch das gesamte vorhandene Wissen aufgebaut wird. In diesem Sinn sind Prinzipien (logisch) vor detaillierteren Aussagen der jeweiligen Wissenschaft und haben einen generelleren Anwendungsbereich als spezialisiertere All- oder Einzelaussagen. In der künstlichen Intelligenz erhofft man sich durch die Aufdeckung der Prinzipien von Intelligenz einen leichten und gangbaren Weg zu intelligenten Systemen. In der Kognitionswissenschaft rechnet man damit, durch die Erkenntnis der Prinzipien eine allgemeine Erklärung für das menschliche Denken zu erhalten.

Zunächst soll jedoch das durch natürliche Intelligenz hervorgebrachte *Verhalten* näher betrachtet werden. Alan M. Turing ging in seinem Artikel *Computing machinery and intelligence* 1950 [Tur50] davon aus, daß sich alle relevanten Äußerungen von Intelligenz gleichermaßen zeigen, wenn das intelligente Wesen ausschließlich schriftlich (d.h. über einen Fernschreiber) mit seiner Umwelt kommuniziert.[9] Die folgenden Betrachtungen lehnen sich an diese Annahme an.

[7] Wittgenstein [Wit53] §66 und §67.

[8] Nach Cantors ursprünglicher Fassung des Mengenbegriffs läßt sich 'die Menge aller Mengen, die sich nicht selbst als Element enthalten' definieren. Dies führt jedoch immer zum Widerspruch - ob die genannte Menge sich selbst enthalten solle oder nicht ! Russells Brief an Frege, in dem er Frege die neu entdeckte Mengenantinomie mitteilte, ist in Heijenoort [Hei70] abgedruckt.

[9] Vergleiche auch Kapitel 9 in dem der Turingtest beschrieben ist.

Alle schriftliche Kommunikation läßt sich dargestellt durch entsprechend lange Zeichenketten auffassen. Somit werden bei einem Dialog bestimmte Zeichenketten in abwechselnder Reihenfolge von beiden Partnern geäußert. In der Regel ist die Äußerung eines Partners von den vorhergehenden Äußerungen seines Dialogpartners abhängig. Mithin läßt sich also ein lebenslanger Dialog eines intelligenten Wesens W mit seiner Umwelt wie folgt beschreiben: W äußert eine Zeichenkette Z_{A_1} der Länge l_{A_1}. Daraufhin empfängt W eine Zeichenkette seiner Umwelt Z_{E_1} der Länge l_{E_1}. W äußert daraufhin eine Zeichenkette Z_{A_2}, deren Gestalt und Länge eventuell von Z_{E_1} abhängen usw. Da man von einem intelligenten Wesen erwartet, daß es auch dann eine als *intelligent* empfundene Äußerung Z_{A_2} tätigt, falls Z_{E_1} eine beliebige andere Zeichenkette gewesen wäre, muß W also ein ganzes Repertoire an Reaktionen auf mögliche Äußerungen seiner Umwelt bereit halten. Formal könnte man dies wie folgt beschreiben:
W hält für jede Zeichenkette die es empfangen könnte, eine geeignete Reaktion bereit. Bei einer empfangenen Binärzeichenkette der Länge l_{E_1} muß W für $2^{(l_{E_1})}$ verschiedene Fälle gewappnet sein. Hierbei werden die allermeisten Zeichenkombinationen freilich keine sinnvollen Sätze ergeben. Wie dem auch sei, die dritte Äußerung von W, Z_{A_3} wird abhängig von der ersten und der zweiten empfangenen Zeichenkette Z_{E_1} und Z_{E_2} sein. Angenommen, die Summe der Längen aller empfangenen Zeichenketten während der Lebenszeit von W, $l_{E_{ges}}$ sei höchstens 10^{15}. Ebenso sei die Summe $l_{A_{ges}}$ der Längen aller geäußerten Zeichenketten von W höchstens 10^{12}. Dann muß W also höchstens auf $2^{(10^{15})}$ verschiedene Äußerungen mit eigenen Äußerungen von einer Gesamtlänge von höchstens 10^{12} Zeichen reagieren können. Dies ist eine sehr grob abgeschätzte obere Grenze. Für die ersten Äußerungen von W müssen erheblich weniger Fälle vorgesehen sein, als für die letzten Äußerungen von W, da die ersten Äußerungen nicht so stark von den Umweltreaktionen abhängen. Die Gesamtheit all dieser potentiellen Reaktionen könnte man in einer einzigen sehr langen Zeichenkette Z_{Int} speichern. In Z_{Int} wäre dann die gesamte Information enthalten, die ein intelligentes Verhalten erfordert. Es könnte für jede der möglichen $2^{(10^{15})}$ verschiedenen Umweltäußerungen jeweils
eine Teilzeichenkette von der Länge 10^{12} nacheinander in Z_{Int} enthalten sein.[10] Somit wäre Z_{Int} von der Länge $2^{(10^{15})} \times 10^{12} \approx 3 \times 10^{(10^{14})} \approx 10^{100\ 000\ 000\ 000\ 000}$.

Es ist ziemlich klar, daß sich eine so lange Zeichenkette nicht physikalisch darstellen läßt - bei einer geschätzten Zahl von Atomen im Weltall von $\approx 10^{80}$. Trotzdem ist es fruchtbar, über diese Zeichenkette zu sprechen. Wie in dem Abschnitt über die algorithmische Informationstheorie gezeigt wurde, hat jede Zeichenkette eine gewisse inhärente Komplexität. Je nach dem wieviel Regelmäßigkeit in Z_{Int} enthalten ist, läßt sich Z_{Int} durch eine eventuell erheblich kürzere Zeichenkette beschreiben, wobei die Beschreibung durch eine universelle Turingmaschine interpretiert wird. Angenommen, die kürzestmögliche Darstellung von Z_{Int}, also die Kolmogoroffkomplexität von Z_{Int}, $K(Z_{Int})$ ist von der Größenordnung 10^{12}. Dann würde ein entsprechendes intelligentes Verhalten einen Gehalt von 10^{12} Bits

[10] Sozusagen würde Z_{Int} eine gigantische Tabelle repräsentieren, in der ich für alle möglichen Wahrnehmungen, die ich nach und nach in meinem Leben machen könnte, ein passendes Verhalten nachschlagen kann.

an algorithmischer Information erfordern.
Weiterhin sei angenommen, in der KI-Forschung bzw. in der Kognitionswissenschaft würden bestimmte Techniken entdeckt, die als die *Prinzipien* von Intelligenz bzw. als die *Prinzipien* des Denkens gelten sollen. Eine präzise Beschreibung dieser Techniken wird nicht zu lang sein. Beispielsweise wird die Beschreibung höchstens von der Länge 10^6 Bits sein dürfen. Erfordern die Techniken eine längere Beschreibung, so wird ein Mensch sie in ihrer Gesamtheit kaum noch übersehen und verstehen können. Sie werden daher auch schwerlich als die *Prinzipien* des Denkens gelten können. Mit diesen Prinzipien wären also höchstens 10^6 Bits der angenommenen 10^{12} Bits, die für ein intelligentes Verhalten notwendig sind bekannt. Somit wären für eine vollständige Beschreibung von intelligentem Verhalten noch die verbleibenden $10^{12} - 10^6 \approx 10^{12}$ Bits an algorithmischer Information erforderlich. Das heißt, der Anteil der *Prinzipien* an der insgesamt erforderlichen algorithmischen Information würde in jedem Fall von einer vernachlässigbaren Größe sein. Mit anderen Worten würden bei der Konstruktion von intelligenten Systemen die entdeckten *Prinzipien* nur einen vernachlässigbaren Teil der Arbeit erledigen - verglichen damit, daß ohne Kenntnis dieser Prinzipien unmittelbar eine *universelle Turingmaschine* programmiert werden müsste.[11] Dennoch stellt sich die Frage, wodurch sich Prinzipien von Intelligenzleistungen auszeichnen - wenn man davon überhaupt sprechen kann.

5.3 Der Allgemeinbegriff von 'Intelligenz'

Was haben *Prinzipien* der Intelligenz oder Prinzipien kognitiver Systeme mit Allgemeinbegriffen, mit Universalien zu tun ?
Einerseits stellt sich die Frage nach der Gesamtheit der Intelligenzphänomene, wenn eine umfassende Theorie entwickelt werden soll. Andererseits zählt auch die Klärung gerade dieser Frage selbst zu den Intelligenzphänomenen, und zwar in zweifacher Weise: Erstens: im täglichen Umgang müssen Phänomene als Intelligenzphänomene erkannt werden - beispielsweise beim Umgang mit bzw. der Einschätzung von Mitmenschen. Zweitens: bei der wissenschaftlichen Auseinandersetzung mit dem Intelligenzphänomen muß die Gesamtheit der zu untersuchenden Phänomene erfaßt werden.
Sowohl in der künstlichen Intelligenz als auch in der Kognitionswissenschaft geht es darum, Prozesse zu beschreiben, die Kognitionen zugrunde liegen, bzw. die geeignet sind, die Ergebnisse kognitiver Prozesse zu erzielen. Zur präzisen, operationalen Beschreibung von Prozessen, bei denen die Ergebnisse *effektiv berechnet* werden können, steht uns nur der Begriff des *Algorithmus* z.B. in Form einer Turingmaschine zur Verfügung.[12]
Will man nun Prinzipien kognitiver oder intelligenter Prozesse bestimmen, so hat man dabei die folgende Aufgabe:

[11] Man kann sich dabei vorstellen, daß vielleicht nicht jedes Bit der notwendigen algorithmischen Information den gleichen Schwierigkeitsgrad seiner Entwicklung hat. Vielleicht sind die Prinzipien besonders schwierig zu entwickeln. Und - wenn man erst einmal die Prinzipien gefunden hat - der Rest sehr viel leichter. In Abschnitt 7.4 und Kapitel 10 wird diese Fragestellung erörtert.

[12] Die Churchsche These behauptet, daß auch nie eine ausdrucksstärkere Beschreibungsmöglichkeit gefunden werden wird. Vergleiche dazu Kapitel 3.

Zum einen soll die Beschreibung der Prinzipien kurz sein - anderenfalls ist sie unübersichtlich und kaum verstehbar. Weiterhin sollen die Prinzipien ja so etwas wie die 'Kernalgorithmen' der kognitiven Prozesse sein, während zusätzlich erforderliche Beschreibungen weniger wichtiges 'Beiwerk' sind.
Das heißt, die Aufgabe besteht darin, aus einer der möglichen sehr umfangreichen Beschreibungen kognitiver Prozesse - oder zumindest aus einer Beschreibung der *Ergebnisse* kognitiver Prozesse - einen kleinen Teil der Beschreibung als den 'Kern der Algorithmen' auszusuchen und als *Prinzipien* zu bestimmen.
Bei einer sehr langen Zeichenkette, in welcher die geforderte Beschreibung der kognitiven Prozesse codiert ist, gibt es eine ungeheuer große Zahl von Möglichkeiten,[13] einen kleinen Teil davon auszusuchen und als Prinzipien zu deklarieren. Man könnte beispielsweise die 'Prinzipien' in einer Spezialmaschine von der 'Software' in die 'Hardware' verschieben, ohne dabei die resultierende Funktionalität des Gesamtsystems zu verändern.
Genau dies ist ja die Idee Pylyshyns für eine Grundlegung der Kognitionswissenschaft.[14] Diese Spezialmaschine entspricht bei den rein theoretischen Betrachtungen, der Definition einer entsprechenden - eventuell universellen - Turingmaschine, die nur noch den verbleibenden Teil der ursprünglichen Zeichenkette als Programm benötigt. Bei der Ausführung dieses Programms wird dann genau die ursprünglich intendierte Funktionalität erreicht.
Wie dem auch sei, es gibt jedenfalls keine *logischen*, keine zwingenden Gründe, gerade irgendeinen *bestimmten* Teil der ursprünglichen Zeichenkette als 'Prinzipien' zu akzeptieren.
Vielmehr können nur *universalistische* Vorstellungen von 'Intelligenz' oder 'kognitiven Prozessen' als Gründe für die Bestimmung bestimmter 'Kernalgorithmen' als die Prinzipien herhalten. Denn die Bestimmung von 'Kernalgorithmen' setzt eine umfassende Betrachtung des Intelligenzphänomens voraus - schließlich sollen sie ja zur Erklärung *aller* denkbaren Intelligenzphänomene brauchbar sein.

> Es ist also eine Vorstellung von Intelligenz oder von kognitiven Prozessen erforderlich, die eine *unendliche* Menge von 'erwünschten' Ein-/Ausgabewertepaaren identifiziert. Dadurch erst kann sich ein bestimmter Teil der algorithmischen Information als 'Prinzipien' herauskristallisieren. Sonst - bei einer nur endlichen Menge - könnte man keinen klaren Unterschied zwischen einem systematischen Verfahren, das die Intelligenzleistungen hervorbringt, und dem Nachschlagen in einer sehr großen, aber endlichen Tabelle machen.[15]

Dies entspricht der Forderung bei der Universalie von beispielsweise *Tisch* Kriterien zu bestimmen, die es ermöglichen, Tische von anderen Objekten zu unterscheiden. Bei dem

[13] Vergleiche Abschnitt 3.4.

[14] Siehe Abschnitt 4.1 für eine detaillierte Beschreibung von Pylyshyns Ansatz. Die genannte Spezialmaschine würde Pylyshyns 'funktionaler Architektur' entsprechen.

[15] Dies scheint zunächst in gewissem Widerspruch zu den in Abschnitt 5.2 angestellten Betrachtungen zu den faktisch feststellbaren Intelligenzleistungen zu stehen. Aber diese Problematik soll ja gerade zum Ausdruck kommen.

Begriff *Intelligenz* ist der Sachverhalt abstrakter - jedoch besteht auch hier das Problem, woher die universelle Vorstellung kommt und wie sie begründet werden kann. In der Tat gehört die Subsumption einer individuellen Inetlligenzleistung unter die Intelligenzphänomene ebenso wie die universelle Vorstellung und deren Begründung zu dem zu charakterisierenden Phänomenbereich.
Wenn gefordert wird, den Gegenstandsbereich der künstlichen Intelligenz scharf abzugrenzen, so ist es nicht nur so, daß - wie sonst üblich - verschiedene kognitive Leistungen als *intelligent* erkannt werden sollen. Hingegen wird auch noch vom erkennenden Subjekt verlangt, daß es selbst seinen Begriff von *Intelligenz* soweit vollständig operationalisiert, daß ein Formalismus - eine Turingmaschine - erkennen kann, ob eine Leistung als *intelligent* zu bezeichnen ist. Somit muß also die gesamte Begriffsextension expliziert werden.
Da sich jedoch keine unendliche Zahl von Instanzen faktisch aufzählen läßt, ja nicht einmal eine größere Anzahl davon, so bleibt nur eine allgemeine Beschreibung, die bestimmte Erkennungsmerkmale nennt; beispielsweise die als 'Prinzipien' behaupteten Strukturen !

5.4 Universalientheorien und der Allgemeinbegriff 'Intelligenz'

Wo liegt der Ursprung einer allgemeinen Vorstellung von Intelligenz ? Und wie kann eine solche gegebenfalls begründet werden ? Im folgenden werden die Instanzen der allgemeinen Vorstellung von Intelligenz als je eine Verhaltensreaktion in einzelnen Situationen aufgefaßt.
Platons Theorie enstprechend hat jeder Mensch, jeder Wissenschaftler der künstlichen Intelligenz oder der Kognitionswissenschaft, intellektuellen Zugang zu dem Allgemeinbegriff von Intelligenz. Dadurch ist er auch in der Lage, von einer einzelnen oder einer kleinen Menge von Intelligenzreaktionen auszugehen und zu verallgemeinern. Durch Platons Form von Intelligenz wird die 'Richtung' der Verallgemeinerung geleitet. Die Beispiele dienen dem Wissenschaftler dabei nur zur 'Erinnerung' an die Form 'Intelligenz', die er vor seiner Geburt bei der platonischen Ideenschau bereits kennenlernte.
Beispielsweise hatten Ende der 50er Jahre Newell & Simon [NS63] ihren *General Problem Solver* (GPS) entwickelt, der an einigen Beispielproblemen, wie etwa den Türmen von Hanoi, getestet wurde. Newell & Simon generalisierten von der Lösung dieser Beispielprobleme auf die Simulation menschlicher Intelligenz. Sie schrieben in einem Bericht:

> Wir haben gerade zu erkennen begonnen, wie wir Computer benützen können, um Probleme zu lösen, für die wir keine systematischen und effizienten Algorithmen besitzen. Und zumindest in einem begrenzten Gebiet wissen wir heute nicht nur, wie man Computer programmieren muß, damit sie erfolgreich Probleme lösen können; wir wissen auch, wie man Computer programmieren muß, damit sie diese Fähigkeiten *erlernen*.
>
> Kurz gesagt, wir kennen nun die Elemente einer Theorie des heuristischen (im Gegensatz zum algorithmischen) Problemlösens; und mit dieser Theorie

> können wir sowohl heuristische Prozesse im Menschen verstehen als sie auch auf Digitalcomputern simulieren. Intuition, Einsicht und Lernen sind nicht länger ausschließlicher Besitz des Menschen: Jedem großen Hochgeschwindigkeitscomputer können sie einprogrammiert werden.[16]

Heute weiß man, daß der GPS nur für eine sehr beschränkte Klasse von Problemen wirklich geeignet ist. Aber Ende der 50er Jahre waren Newell & Simon offensichtlich von der Allgemeinheit der in GPS programmierten Prinzipien überzeugt.
Daß Newell & Simon annahmen, daß alle oder zumindest ein erheblicher Teil von Intelligenzleistungen durch die Prinzipien des GPS erklärt und erzeugt werden können, ist somit durch eine entsprechende platonische Idee des Allgemeinbegriffs von Intelligenz zu erklären.
Sie hätten auch in andere Richtungen ihre erfolgreichen Versuche generalisieren können, z.B. auf alle Rätselaufgaben, wie Türme von Hanoi, 15-Puzzle, Rubiks Cube etc., oder aber auf alle Spiele, einschließlich Schach und Go etc. etc.
Erst durch ihre weitgehende Generalisierung, erst durch die Annahme ihre Beispiele seien repräsentativ für allgemeine Intelligenzleistungen, mithin erst durch eine entsprechende Vorstellung vom Allgemeinbegriff 'Intelligenz' entstand der Anspruch, die allgemeinen Prinzipien von Intuition, Einsicht und Lernen - welches ganz zentrale Fähigkeiten intelligenter Wesen sind - erkannt zu haben.

Eine andere Möglichkeit der Erklärung, wie Wissenschaftler zu der Annahme kommen, die Prinzipien des Denkens oder von Intelligenz entdeckt zu haben, ist die Abstraktion im Sinne der Aristotelischen Universalientheorie.
Diese Erklärung erscheint insbesondere dann plausibel, wenn von introspektiven Erfahrungen ausgegangen wird. Dann werden Gedankengänge beobachtet und von ihren spezifischen Inhalten abgesehen, abstrahiert. Dies ist beispielsweise die Vorgehensweise in der Entwicklung der Logik gewesen. Aristoteles entwickelte Syllogismen, bei denen von einem konkreten Inhalt abstrahiert wurde, so daß nur noch die inhaltsleere Schlußform übrig blieb:[17]

Beispiel:

Alle Menschen sind sterblich.	**Verallgemeinerung:**	$\forall x\ P(x) \rightarrow Q(x).$
Alle Griechen sind Menschen.		$\forall x\ R(x) \rightarrow P(x).$
Alle Griechen sind sterblich.		$\forall x\ R(x) \rightarrow Q(x).$

Hierbei ist allerdings zu bemerken, daß es bei der Aristotelischen Abstraktion nicht klar ist, ob daraus die Prinzipien von Intelligenzleistungen gewonnen werden. Denn es wird dabei von einer *Vorgehensweise* ausgegangen und verallgemeinert - und nicht, wie oben beschrieben, von den Intelligenz*leistungen*. Letzteres ist aber für eine künstliche Intelligenz

[16] Übersetzt aus: Simon & Newell [SN58] in Dreyfus [DD87] S. 25-26.
[17] Siehe z.B. Prantl [Pra55] für eine Darstellung der historischen Entwicklung der Logik.

das Entscheidende, die primär daran interessiert ist, Intelligenzleistungen auf irgendeine Weise zu erzeugen. Demgegenüber würde die genannte Aristotelische Abstraktion als Simulation einer bestimmten Vorgehensweise stehen, die zudem nur zu einem Teil zur Hervorbringung menschlicher Intelligenzleistungen *beiträgt*.

Im Gegensatz zu diesen beiden auf Realismustheorien basierenden Erklärungen für die Generalisierungen von einzelnen Intelligenzleistungen oder Denkprozessen auf die 'allgemeine Form' von Intelligenz bzw. des Denkens stehen die Ähnlichkeitstheorien, insbesondere Wittgensteins Begriff der *Familienähnlichkeit*: Die Familienähnlichkeiten lassen sich *nicht* durch eine einfache Regel beschreiben.[18] So wird auch eine *algorithmische* Beschreibung komplizierter sein müssen.
Bei Anwendung von Wittgensteins Überlegungen auf den Allgemeinbegriff *Intelligenz* kommt man ebenfalls zu dem Schluß, daß die Abgrenzung des Intelligenzbegriffs sich nicht durch die Angabe einiger weniger Charakteristika durchführen läßt.
Will man Intelligenzleistungen charakterisieren, so wird mal das logisch stringente Denken im Vordergrund stehen, mal das schnelle Kopfrechnen, dann das rasche Erkennen von visuellen und akustischen Reizen, das schnelle und sichere Erinnern an selten benutzte Kenntnisse, dann der kreative Gedanke, das geeignete induktive Schließen, dann das angemessene Verhalten in Situationen, in denen nur unsicheres Wissen zur Verfügung steht, dann das Finden von passenden Analogien oder gar von Metaphern etc. etc.
Das wichtigste jedoch ist, und dies ist in der Tat wenig offensichtlich und eher irreführend, daß die meisten der oben verwendeten Ausdrücke selbst wiederum nur unklar sind und ihrerseits nur durch Familienähnlichkeiten unter ihren Instanzen zu erklären sind. Mithin täuschen die Ausdrücke über die tatsächliche Uneinheitlichkeit der Struktur menschlicher Intelligenz hinweg. Diese Behauptung wird auch durch die Tatsache gestützt, daß die Bemühungen der künstlichen Intelligenz, entsprechende Fähigkeiten in Form eines Computerprogramms operational zu beschreiben, bisher nur von mäßigem Erfolg gekrönt sind.
Wie dem auch sei, Wittgensteins Überlegungen deuten bereits an, daß allein die *Abgrenzung* von intelligenten gegenüber weniger intelligenten Leistungen eine hohe Kolmogoroffkomplexität erfordert. Dies würde sich - allein unter dem Gesichtspunkt der Beschreibungskomplexität - mit den realistischen Universalientheorien kaum vereinbaren lassen. Die realistischen Theorien würden hingegen den Menschen dazu verleiten, eine Grenze zu ziehen, wo noch keine Grenze gezogen ist, um mit Wittgensteins Worten zu sprechen.

Vergleiche hierzu Wittgenstein:

> ... Denn ich *kann* so dem Begriff $\succ$ Zahl $\prec$ feste Grenzen geben, d.h. das Wort 'Zahl' zur Bezeichnung eines fest begrenzten Begriffs gebrauchen, aber ich kann es auch so gebrauchen, daß der Umfang des Begriffs *nicht* durch eine Grenze abgeschlossen ist. ... Kannst Du die Grenzen angeben ? Nein. Du

[18] Es handelt sich ja nicht nur um einen einzelnen Aspekt, der den subsumierten Objekten gemeinsam ist. Vielmehr sind es mal diese mal jene gemeinsamen Aspekte, die die subsumierten Einzelfälle aufweisen.

> kannst welche *ziehen*: denn es sind noch keine gezogen. (Aber das hat dich noch nie gestört, wenn du das Wort ≻ Spiel ≺ angewendet hast.) [19]

Diese Grenze wird dabei allerdings weniger explizit gezogen. Es werden vielmehr methodologische Annahmen gemacht, die eine Grenze implizieren. Beispielsweise wie es Newell & Simon mit dem GPS taten. Später stellte sich heraus, daß die implizit gezogene Grenze, - nämlich all das, was der GPS kann, gilt als Intelligenzleistungen - mit der allgemeinen Auffassung von Intelligenz in keiner Weise übereinstimmt. Im folgenden Kapitel wird auf diese Problematik noch näher eingegangen.

Wie dem auch sei, so haben sich doch heute in der KI-Forschung unter anderem die bereits in Kapitel 2 skizzierten Teilgebiete etabliert: *Suchen und Problemlösen, Spiele spielen, Automatisches Beweisen, Verstehen natürlicher Sprache, Bildverstehen, Lernen, Expertensysteme* sowie *Konnektionismus und neuronale Netzwerke*.
Jedes dieser Teilgebiete beinhaltet eigene Aufgabenstellungen und man versucht, geeignete Lösungstechniken für die jeweiligen Aufgabenstellungen zu entwickeln. Der Trend geht in der Tat in die Richtung, die Probleme immer feiner zu klassifizieren und für jede dieser Problemklassen spezielle Verfahren zu entwickeln. Bei einer solchen Untergliederung der Techniken stellt sich allerdings die Frage, inwiefern eine solche Vielzahl von Techniken noch die *Prinzipien* von Intelligenz widerspiegeln können.[20]

[19] Wittgenstein [Wit53] §69.

[20] Der starke Trend zur feineren Klassifizierung des Fachgebietes läßt sich nicht nur innerhalb der KI, sondern allgemein in der Informatik in einem weit größeren Maße beobachten, als in anderen Wissenschaften. Mahr [Mah84] bietet eine philospohisch-soziologische Diskussion dazu.

Kapitel 6

Methodologische Probleme der künstlichen Intelligenz

Welche Konsequenzen haben die im vorhergehenden Kapitel aufgezeigten Probleme für die Forschungsmethoden im Bereich der KI und der Kognitionswissenschaft ? Smolensky [Smo88, Smo90] und anderen folgend ist zumindest ein Ziel der Kognitionswissenschaft, die *Prinzipien*, die den kognitiven Prozessen zugrunde liegen, zu entdecken. Wie im vorhergehenden Kapitel ausgeführt wurde, ist es aber keineswegs klar, was solche *Prinzipien* gegenüber anderen Faktoren auszeichnet. Es wurde bereits darauf hingewiesen, daß eine Turingmaschine ausreicht, um intelligentes Verhalten hervorzubringen. Wenn die Annahme zutrifft, daß dafür eine erhebliche algorithmische Information erforderlich ist, so wirft dies allerdings generelle Probleme für die Erkenntnismöglichkeiten in der KI und der Kognitionswissenschaft auf.
Im ersten Abschnitt wird zunächst argumentiert, daß sich die Prinzipien der Intelligenz nicht auf 'natürliche' Weise im Rahmen der Forschungsbemühungen in der KI bzw. der Kognitionswissenschaft herausbilden werden. Der zweite Abschnitt diskutiert, wodurch sich *Prinzipien* der Intelligenz überhaupt auszeichnen können. In Hoffmann [Hof91c] wurden die Grundgedanken von mir erstmals veröffentlicht. Abschnitt 6.3 wird diese Problematik am Beispiel der hypothetischen Entwicklung einer Theorie des maschinellen Lernens[1] erläutert.

6.1 Die Begründung von Prinzipien

In diesem Abschnitt soll dafür argumentiert werden, daß jeglicher Grund, der etwas als die Prinzipien der Intelligenz auszeichnen könnte, auf eine **platonische Sichtweise** des Phänomens *Intelligenz* zurückgeführt werden kann und weiterhin, daß dies immer in eine Sackgasse weisen muß !

[1] Eine empirische Theorie des *menschlichen* Lernens muß dabei allerdings mit ganz ähnlichen Schwierigkeiten kämpfen. In Arnold et al. [AEM87] Seite 1258-1266 findet sich ein Überblick über psychologische Lerntheorien, die noch weit davon entfernt sind, menschliches Lernverhalten in praktischen Lernsituationen zu erklären oder gar vorherzusagen.

Für die folgenden Überlegungen wird - in Anlehnung an Ryle [Ryl49] - Intelligenz weiterhin als eine Fähigkeit betrachtet - d.h. Intelligenz als eine gewisse Disposition auf Reize mit entsprechendem Verhalten zu reagieren. Somit könnte man auch sagen, Intelligenz löst das Problem auf Reize in entsprechender Weise zu reagieren. Diese Betrachtungsweise schließt für eine deskriptive Theorie nicht aus, auf intentionale Zustände bezugzunehmen oder Reize intentional zu interpretieren. Das genannte Problem auf Reize zu reagieren, ist dabei eigentlich eine Klasse von Probleminstanzen. Eine Probleminstanz besteht darin, auf *spezifische* Reize[2] geeignet zu reagieren. Wie bereits erläutert wurde, sind Algorithmen *allgemeine* Verfahren. Das heißt, sie lösen eine große (unendliche) Klasse von Probleminstanzen eines bestimmten Problemtyps. Was sind aber die Problemtypen, die *Intelligenz* in der Lage ist, zu lösen ? Was sind die Problemtypen, die *Denken* imstande ist, zu lösen ? in der gegenwärtigen Forschungslandschaft der KI haben sich unter anderem die in Kapitel 2 skizzierten Teilgebiete der künstlichen Intelligenz etabliert.

Mögliche Gründe für gerade diese Unterteilung könnten sein:

- Jedes Gebiet hat eigene, für sich typische Probleme.
- Jedes Gebiet hat seine eigenen, nur dort wichtigen Techniken.
- Jedes Gebiet hat seine Anhänger, die es durch ihre Forschungsanstrengungen formen. Dabei könnten auch Gesichtspunkte der Projektfinanzierung wesentlichen Einfluß auf die Einteilung nehmen.

Ein wichtiger Gesichtspunkt, aufgrund dessen zumindest auch die gegenwärtige Unterteilung vorliegt, ist sicherlich, daß jedes Gebiet eigene Techniken hat. Hierbei sind trotzdem bestimmte Gebiete grundlegender als andere. Beispielsweise finden Techniken aus dem Gebiet *Suchen und Problemlösen* nicht nur beim Probleme lösen, sondern auch auf dem Gebiet des maschinellen Lernens, des Bildverstehens oder des automatischen Beweisens wichtige Anwendungen. Ähnlich ist es mit dem Gebiet des automatischen Beweisens und des maschinellen Lernens. Wie dem auch sei, die Probleme, die in den einzelnen Teilgebieten betrachtet werden, sind oft recht spezieller Natur.
Zum Beispiel untersucht man auf dem Gebiet *Suchen und Problemlösen* mit welchen Verfahren am effektivsten in großen kombinatorischen Räumen ein lokales oder globales Optimum gefunden werden kann; dabei sind Bewertungsfunktionen angegeben, die topologisch nahegelegenen Punkten einen ähnlichen Wert geben *sollen*, d.h. sie tun es nicht in allen Fällen, aber immerhin in einer großen Zahl der Fälle. Die Topologie über dem Suchraum wird dabei von einem zugrunde liegenden Suchverfahren definiert.
Inwiefern solche Probleme als die Probleme angesehen werden können, die *typischerweise* von intelligenten Wesen gelöst werden, ist schwer zu sagen. In ihrer expliziten Formulierung werden sie sicherlich nicht im Alltag bearbeitet. Doch kommt man nicht umhin,

[2] Alle Reize, mit denen das System bereits früher konfrontiert wurde, sind hierbei miteinzubeziehen. Dadurch wird also auch ein mögliches Anpassungs- bzw. Lernverhalten aus früheren Reizen mitberücksichtigt.

festzustellen, daß gerade derartige Suchprobleme sehr wahrscheinlich in fast allen Prozessen des menschlichen Denkens mitenthalten sind - auch wenn es noch unbekannt ist, wie das menschliche Denken im Einzelnen vor sich geht.
Wie dem auch sei, für jede *endliche* Problemklasse, von der angenommen wird, sie werde typischerweise von Intelligenz gelöst, gibt es Algorithmen, die *alle* Probleminstanzen dieser Klasse lösen. So gibt es beispielsweise für jede endliche Zahl von prädikatenlogischen Formeln einen Algorithmus, der entscheidet, ob sie innerhalb eines gegebenen Axiomensystems gelten. Fordert man hingegen einen Algorithmus, der beliebige prädikatenlogische Formeln entscheidet, so wurde gezeigt, daß es einen solchen Algorithmus nicht gibt.[3] Weiterhin läßt sich beobachten, daß es für eine endliche Zahl von Formeln nicht nur *einen* korrekten Algorithmus gibt. Es gibt sogar beliebig viele verschiedene Algorithmen, die auch in ihrer Grundstruktur sehr verschieden sein können.
Diese Unterschiede können sich sogar in ihrem Ein-/Ausgabeverhalten zeigen. Dabei können die verschiedenen Algorithmen für Eingaben, die *nicht* zu der vorgesehenen endlichen Zahl von Formeln gehören, ganz unterschiedliche Ausgabeverhalten zeigen. Diese Unterschiede wären für die intendierte Aufgabenstellung - die genannte endliche Zahl von prädikatenlogischen Formeln zu entscheiden - nicht von Bedeutung, weil sich ihr potentielles - sehr unterschiedliches - Verhalten bei den intendierten Eingaben nie zeigen würde. Man kann sich die Ausgaben eines geeigneten Algorithmus für die intendierte endliche Zahl von Formeln, die es korrekt zu entscheiden gilt, auch als den Anfang einer unendlichen Zeichenkette vorstellen. Abgekürzt $Z = AR$, wobei Z die unendliche Zeichenkette bezeichnet, A den endlichen Anfang von Z und R den unendlichen Rest von Z, wobei in Z die gewünschten Ausgaben auf die verschiedenen Eingaben in einer festgelegten Reihenfolge codiert sind. Der Anfang A der Zeichenkette ist festgelegt - jedes Zeichen muß der korrekten Antwort auf eine der intendierten Formeln beschreiben. Der Rest der Zeichenkette R jedoch kann beliebig aussehen. Aus der Sicht der algorithmischen Informationstheorie kann man sagen, daß sich eine solche Zeichenkette *immer* zumindest auf die Länge von A komprimieren läßt.
Das Beispiel der Entscheidbarkeit prädikatenlogischer Formeln scheint vielleicht kein gutes Beispiel zu sein, da immerhin für eingeschränkte Bereiche Algorithmen entwickelt wurden, die tatsächlich für eine unendliche Zahl von Formeln die richtige Antwort bestimmen. Zum Beispiel für die Aussagenlogik, Aristoteles Klassenlogik oder für die prädikatenlogischen Formeln der Presburger-Arithmetik.[4] In diesen Fällen gäbe es keinen Rest R, es würde $Z = A$ gelten, da ein korrektes Berechnungsergebnis für alle (unendlich vielen) Eingabewerte gefordert ist. Für diese eingeschränkten Bereiche gibt es also sogar eine sehr kurze Beschreibungen von Z, das heißt, Z könnte sehr stark komprimiert werden. Der Bereich des logischen Schließens wurde spätestens seit der griechischen Antike als eine geistige Tätigkeit angesehen, die man bemüht war zu formalisieren. Mithin zweifelte eigentlich niemand daran, daß es *Prinzipien* des logischen Schließens gibt, die eben nur noch in einer

[3] Dies geht aus dem Beweis der Semientscheidbarkeit der Prädikatenlogik erster Stufe hervor. Diese wurde zuerst von Church [Chu36a] und von Turing [Tur37] bewiesen.

[4] Die Presburger-Arithmetik beinhaltet den Bereich der natürlichen Zahlen mit Addition und der Größer-Relation ($>$).

entsprechenden Formalisierung fixiert werden müssten.
Einer solchen Fixierung der Prinzipien entspricht die extreme Komprimierung von der unendlich langen Zeichenkette Z auf eine endliche, kurze Zeichenkette. Diese komprimierte Beschreibung von Z repräsentiert dabei einen Schlußkalkül für den entsprechenden Bereich. Einige Ergebnisse der heutigen Informatik legen allerdings gewisse Zweifel an der Annahme nahe, daß der Bereich des logischen Schließens zu den Prinzipien des menschlichen Denkens gehört.
Beispielsweise bewiesen Fischer und Rabin 1974 [FR74], daß für jeden Ableitungsalgorithmus und für jede endliche Axiomatisierung der Presburger-Arithmetik der Beweis von (*schwierigen*) Formeln sehr lang wird. Genauer gesagt wächst die Zahl der erforderlichen Beweisschritte hyperexponentiell mit der Länge der Formel, sie ist also von der Größenordnung 2^{2^n} bei einer Formellänge von n.
Ein anderes Ergebnis von Cook 1971 [Coo71], der Beweis der sogenannten **NP**-Vollständigkeit des Erfüllbarkeitsproblems aussagenlogischer Formelmengen deutet in eine ähnliche Richtung. Nach einer bisher noch unbewiesenen, aber von den allermeisten der führenden theoretischen Informatiker angenommenen Hypothese (der sogenannten $\mathbf{P} \neq \mathbf{NP}$-Hypothese) bedeutet das Ergebnis von Cook, daß es *keinen* Algorithmus gibt, der für aussagenlogische Formelmengen ihre Erfüllbarkeit entscheidet und dafür weniger als eine in der Länge der Formelmenge exponentiell wachsende Zahl von Rechenschritten benötigt, obgleich ein Beweis der Erfüllbarkeit erheblich kürzer ist. Die benötigte Zahl von Rechenschritten rührt also nur daher, daß es nicht möglich ist, einen Erfüllbarkeitsbeweis *systematisch* mit weniger Rechenschritten zu *finden*, bzw. auszuschließen, daß es einen Solchen gibt.
Mit anderen Worten deuten diese Ergebnisse darauf hin, daß dem menschlichen logischen Schließen andere *Prinzipien* zugrunde liegen, als die ausgearbeiteten formalen Ableitungskalküle. Ja, sie deuten an, daß es gar keine solchen Ableitungskalküle bzw. etwas, das durch ein Ableitungskalkül vertreten werden kann, im menschlichen Denken gibt - auch nicht für Bereiche wie die Aussagenlogik.
Die Überlegung, daß das menschliche Denken vermutlich hochgradig parallel vor sich geht, genügt nicht um die Bewältigung des erforderlichen großen Rechenaufwandes für die Entscheidung aussagenlogischer Formeln zu erklären. Bei einer geschätzten Zahl von 10^{11} Neuronen, kann dies auch nur in einer um einen Faktor von höchstens 10^{11} beschleunigten Verarbeitung resultieren.[5] Diese Konstante von 10^{11} wird bei dem exponentiellen Wachstum der Zahl der Rechenschritte bei größeren Formelmengen rasch 'aufgefressen'. Schließlich muß man wohl davon ausgehen, daß ein Mensch mit Millionen von Aussagen über seine Lebenswelt umgehen kann.

All dies läßt also selbst Zweifel daran rechtfertigen, daß das logische Schließen in dem Sinne, wie es Aristoteles zu formalisieren begann, zu den *Prinzipien* des menschlichen Denkens oder der menschlichen Intelligenz gehört.[6] Worüber es sicherlich keine Zweifel gibt, ist, daß Menschen bei kleinen Formelmengen korrekt logisch schließen können.

[5] Siehe Feldman et al. [FFGL90] Seite 434 für eine Beschreibung des menschlichen Gehirns.

[6] Zu einem ähnlichen Schluß ist auch Frixione [Fri91] gekommen. Vergleiche hierzu auch Levesque [Lev88a].

Mithin sie also für eine endliche, genauer eine kleine Zahl von Formeln korrekte Schlüsse ziehen können. Für diese kleine Zahl von Formelmengen lassen sich allerdings eine Vielzahl von Algorithmen denken, die für diese bestimmten Formelmengen korrekt arbeiten.[7] Je größer die Menge der geforderten richtig zu entscheidenden Formelmengen ist, desto weniger Freiheitsgrade für die Wahl eines geeigneten Algorithmus gibt es.
Dies würde im Extremfall in Ergebnissen münden, wie dem von Kurt Gödel: Daß es überhaupt keinen Algorithmus für das allgemeine Entscheidungsproblem der Prädikatenlogik gibt.[8]
Durch die Erweiterung der Klasse von Problemen, die beispielsweise von einem Deduktionsalgorithmus gelöst werden sollen, wird die Wahl der möglichen Algorithmen zunehmend eingeschränkt. In diesem Sinn ist eine Idealisierung von bestimmten (*endlichen*) Problemklassen, die Menschen zweifelsfrei durch ihr Denken lösen können, zu einer großen (eventuell unendlichen) Problemklasse dafür verantwortlich, daß sich bestimmte Verfahren - bzw. eine bestimmte Klasse von Verfahren - als die einzig zulässigen herauskristallisieren. Erst durch eine solche Idealisierung von bestimmten Problemklassen lassen sich auch einige Teile der algorithmischen Information, die für das Hervorbringen intelligenten Verhaltens erforderlich ist, als *Prinzipien* gegenüber dem übrigen Teil der algorithmischen Information auszeichnen.

Eine mögliche Trennung könnte beispielsweise wie folgt aussehen:

- Prinzipien:

  Ableitungskalkül für die Prädikatenlogik erster Stufe.

- Rest der algorithmischen Information:

  Aussagen über die Lebenswelt des Individuums.

Wie oben ausgeführt, scheint eine solche Unterteilung allerdings nicht angemessen zu sein. Aber wie sollte stattdessen eine angemessene Unterteilung aussehen ?
Welche Unterteilung auch vorgenommen wird, das was dann als *Prinzipien* bezeichnet wird, wird eine bestimmte Klasse von Problemen lösen können. Wahrscheinlich werden die Prinzipien so allgemein sein, daß eine *unendliche* Problemklasse von den Prinzipien 'gelöst' werden kann. Diese *unendliche* Problemklasse kann unmöglich aus der Erfahrung stammen. Sie kann vielmehr nur eine Extrapolation von einer endlichen Zahl von Probleminstanzen sein, die aus der Erfahrung stammen können.

[7] Man sollte sich hierbei vor Augen halten, daß viele Menschen, die nicht im logischen Denken geschult sind, bereits bei sehr einfachen Aufgaben Fehler machen. Dies kann als Hinweis darauf gedeutet werden, daß das praktische menschliche Denken in der Tat einen ganz anderen Weg beschreitet, als die Regeln eines Schlußkalküls abzuarbeiten.

[8] Lucas [Luc61, Luc70], Jacquette [Jac87], Penrose [Pen89] und andere haben darauf basierend argumentiert, daß das menschliche Denken nicht mechanisierbar bzw. algorithmisch beschreibbar ist. Gegen eine solche Argumentation wurden allerdings auch starke Einwände unter anderem von Benacerraf [Ben67], Webb [Web68, Web80] und Slezak [Sle87] erhoben. Siehe Kapitel 9 für eine Diskussion der Thematik.

Wohingehend die Extrapolation durchgeführt wird, kann damit nur durch *platonische Ideen* begründet werden. Daß dies problematisch ist - auch für die Praxis der Forschung in der künstlichen Intelligenz - zeigen unter anderem die bisherigen Mißerfolge.
Beispielsweise geht der Trend im Bereich der Wissensrepräsentation und der Deduktionssysteme mittlerweile dahin, Formalismen wie die Prädikatenlogik deutlich einzuschränken, so daß nicht nur vollständige und korrekte Ableitungskalküle existieren, sondern auch effiziente Ableitungsverfahren. Es werden Überlegungen angestellt, die Vollständigkeit aufzugeben, um effiziente Verfahren zu erhalten. Man versucht sogar aus Komplexitätsgründen auf Konsistenz im großen Rahmen zu verzichten, um aus inkonsistenten Datenbasen trotzdem noch einigermaßen nützliche Schlüsse zu ziehen.[9]
Der Leitgedanke bei den Deduktionssystemen, mächtige Formalismen für unendliche Objektbereiche und beliebig große Formelmengen möglichst mit vollständigen und korrekten Ableitungsverfahren zu entwickeln, hat insofern den Erfolg in der künstlichen Intelligenz sicher nicht beflügelt, sondern eher behindert !
Wie in Abschnitt 4.2.2 bereits erwähnt, sind derzeit neue Ansätze in der künstlichen Intelligenz bzw. der Kognitionswissenschaft populär geworden, deren Vertreter sich die Entdeckung der allgemeinen Prinzipien von Intelligenz bzw. menschlicher Kognitionen versprechen.

6.2 Methodologischer Zirkel bei der Suche nach einfachen Prinzipien

Zum einen ist klar, daß eine formale Symbolmanipulation im Sinne der universellen Turingmaschine zumindest a posteriori ausreicht, um Intelligenzleistungen zu erklären. Smolensky [Smo88] und andere erhoffen sich jedoch aus der Postulierung einer subsymbolischen Verarbeitungsebene Erkenntnisse für die Kognitionswissenschaft. Welcher Art können solche Erkenntnisse sein ? Im Hinblick auf die Komplexität der betrachteten kognitiven Prozesse wurde in Hoffmann [Hof90b] gezeigt, daß allein die Beschreibung von neuronalen bzw. subsymbolischen Informationsverarbeitungsmodellen *mindestens die gleiche* algorithmische Information erfordert, wie die Beschreibung der kognitiven Phänomene auf einer symbolischen Ebene. In Hoffmann [Hof90b] wird gezeigt, daß die *genaue Beschreibung* eines konnektionistischen Informationsverarbeitungsmodells *mindestens* von der gleichen Komplexität im Sinne der algorithmischen Informationstheorie ist, wie die von den beschriebenen Modellen berechnete Ein-/Ausgabefunktion. Dies gilt gleichermaßen für lernende konnektionistische Modelle. Bei den konnektionistischen Modellen kommt also zu der ohnehin erforderlichen Beschreibungskomplexität noch die außerordentlich schwierige Analyse der Netzwerkdynamik[10] hinzu, um ein Verständnis für die kognitiven Prozesse zu ermöglichen.
Man könnte an eine mehrschichtige Betrachtung kognitiver Phänomene denken, wie sie etwa von Dennett [Den71, Den87] oder Pylyshyn [Pyl84] vorgeschlagen wurde. Dadurch

[9] Siehe z.B. Brachman [Bra90] oder Bauval & Cholvy [BC91].
[10] Siehe z.B. Papert [Pap88].

würde man zwar sehen, daß kognitive Prozesse durch subsymbolische Prozesse konstituiert werden können, jedoch wird man den Ablauf der kognitiven Prozesse selbst nicht besser beschreiben und vorhersagen können. Gleich welche Betrachtungsebene man wählt, die erforderliche Beschreibungskomplexität, um kognitive Prozesse *präzise* zu beschreiben, kann höchstens größer - nicht aber kleiner werden, als auf einer symbolischen Ebene und einer zugrundeliegenden universellen Turingmaschine.[11]

Aufgrund der angenommenen hohen erforderlichen algorithmischen Information für die Beschreibung intelligenten Verhaltens, liegt damit das folgende methodologische Problem vor:
In empirischen Wissenschaften wird in der Regel ein kleiner überschaubarer Untersuchungsrahmen geschaffen. Innerhalb dieses Rahmens können dann Versuche durchgeführt werden, um auf gesetzesartige Zusammenhänge zwischen einzelnen Phänomenen des Forschungsgebietes schließen zu können. Sobald solch gesetzesartige Zusammenhänge formuliert wurden, können diese in erweiterten Versuchsrahmen geprüft und gegebenenfalls falsifiziert werden.[12] Bei den erweiterten Versuchsrahmen wirken mehr Einflußfaktoren auf die hypothetisch gesetzesartig verbundenen Phänomene ein. Es ist dabei allerdings eine essentielle Voraussetzung, daß die hinzukommenden Einflußfaktoren den formulierten gesetzesartigen Zusammenhang nicht stören. Ansonsten würden sich die innerhalb des ersten überschaubaren Untersuchungsrahmens formulierten gesetzesartigen Zusammenhänge als unzutreffend herausstellen.[13]
Mit anderen Worten besteht in den empirischen Wissenschaften die typische Herangehensweise darin, zunächst kleine Untersuchungsbereiche abzugrenzen, um innerhalb dieser gesetzesartige Zusammenhänge zu erkennen, die ansonsten bei zahlreichen Störeinflüssen schwieriger herauszukristallisieren wären.
Die gleiche Herangehensweise wird in der Regel auch auf dem Gebiet der KI und der Kognitionswissenschaft vorgeschlagen bzw. praktiziert. Dabei werden Algorithmen entworfen und für kleine Bereiche (sogenannte Spielzeugwelten) ausprobiert. Wenn die Algorithmen in diesen Welten erfolgreich eingesetzt werden konnten, so erhofft man sich eine Übertragbarkeit auf größere Bereiche von praktischer Bedeutung. Im folgenden soll an dem Fallbeispiel des maschinellen Lernens erläutert werden, daß ein solches Vorgehen in der KI und der Kognitionswissenschaft besondere Probleme birgt.[14]
Wenn intelligentes Verhalten tatsächlich viel algorithmische Information erfordert - und

[11] Nach dem Invarianztheorem der algorithmischen Informationstheorie ist das Verhalten ein kürzestmögliches Programm für eine gegebene universelle Turingmaschine auf keine Weise kürzer beschreibbar, auch nicht z.B. als Netzwerk von Recheneinheiten.

[12] J. S. Mill [Mil43] entwickelte in Anlehnung an Überlegungen F. Bacons seine sogenannte *Eliminationsmethode*. Bei dieser Methode werden die Gemeinsamkeiten von untersuchten Einzelfällen durch weitere Einzelfälle reduziert, bis eine eindeutig erscheinende Ursache-Wirkungsbeziehung übrig bleibt. Dies enthält eine Reihe von problematischen Voraussetzungen, auf die hier jedoch nicht näher eingegangen werden soll.

[13] Siehe beispielsweise Cochran [CC50], Fisher [Fis60] oder Plutchik [Plu74].

[14] Die Probleme, die beim maschinellen Lernen auftreten, betreffen ebenfalls die Versuche eine präzise deskriptive Theorie des menschlichen Lernens zu entwickeln.

dafür spricht eine ganze Menge, u.a. die bisherigen Versuche in der KI - dann liegt die folgende typische Situation vor:
In einem abgegrenzten Versuchsbereich wird nur ein kleiner Teil der für ein allgemein intelligentes Verhalten erforderlichen algorithmischen Information benötigt.
Angenommen in dem abgegrenzten Bereich hat die Ein-/Ausgabefunktion eines 'intelligenten' Wesens eine Kolmogoroffkomplexität von 1 000 Bits. Wenn ein Programm entwickelt wird, das die geforderte Ein-/Ausgabefunktion realisiert, kann man in aller Regel die Grundidee oder das *Prinzip*, das in dieses Programm eingearbeitet wurde, und die bereichsspezifischen Daten voneinander trennen.[15] Hierbei wird die Trennung zwischen *Prinzip* und dem Rest in der Regel auf den konzeptionellen Ideen der Entwickler basieren, die bereits *vor* der vollständigen Entwicklung vorhanden waren. Eine solche Trennung könnte beispielsweise in dem folgenden quantitativen Verhältnis der algorithmischen Information resultieren:

- Die 'erforschte' Technik des Problembereiches, z.B. Lernen oder Sprachverarbeitung etc. hat eine Kolmogoroffkomplexität von 700 Bits.
- der verbleibende Rest - die bereichsspezifischen Daten - hat eine Kolmogoroffkomplexität von 300 Bits.

Damit liegen Zahlenverhältnisse *Prinzip : Daten* und *Bereichskomplexität : Daten* vor, wie in der untenstehenden Tabelle in der ersten Zeile.
Werden diese Prinzipien auf einen größeren Bereich übertragen, z.B. auf einen Bereich, in dem anstatt 1 000 Bits 100 000 Bits an algorithmischer Information erforderlich sind, so erhält man die Zahlenverhältnisse der zweiten Zeile der Tabelle.

Gesamtkomplexität	Prinzip : Daten	Bereichskomplexität : Daten
1 000 Bits	70 : 30	100 : 30
100 000 Bits	0,7 : 99,3	100 : 99,3

Damit spielen die Prinzipien in dem erweiterten Bereich eine zunehmend geringere Rolle. Je größer der Bereich ist, desto mehr Daten (prozentual von der Gesamtbereichskomplexität) müssen zu den 'erforschten Prinzipien' hinzu kommen, um ein korrekt arbeitendes System in dem erweiterten Bereich zu erhalten. Die Aufgabe der adäquaten Behandlung eines abgegrenzten Bereiches wird also *überproportional* schwieriger mit zunehmender Komplexität des Bereiches.
Daher wird eine lineare Extrapolation von den Schwierigkeiten in einem Versuchsbereich auf einen größeren Bereich, die man bei der Verwendung von entwickelten und getesteten 'Prinzipien' haben wird, typischerweise unzutreffend sein. In der Tat wird die Bedeutung der entwickelten 'Prinzipien' asymptotisch gegen Null gehen. Entsprechendes gilt für deskriptive Theorien kognitiver Prozesse.

[15] Ob dies im Einzelfall möglich ist, hängt von dem spezifischen Programmaufbau ab. Dieser hängt wiederum von dem Programmentwickler ab.

> Der erhoffte Effekt, mittels einer entwickelten 'Technik', die die 'Prinzipien' der menschlichen Herangehensweise oder einer anderen künstlichen und erfolgreichen Strategie repräsentieren soll, in größeren Problembereichen schneller zu einer Lösung zu kommen, wird sich typischerweise *nicht* einstellen.

Diese Schlußfolgerung aus den vorhergehenden Betrachtungen wird in der Tat durch viele Aktivitäten in der KI Forschung bestätigt.
Neben den bereits erwähnten Ansätzen des General Problem Solvers (GPS) von Newell & Simon oder der Entwicklung möglichst mächtiger Logikkalküle, die mittlerweile wieder aufgegeben wurden, gibt es auch einige aktuelle Ansätze, bei denen ihre Vertreter Hoffnung auf die Entdeckung allgemeiner Prinizipien haben:
Dazu zählen die erwähnten subsymbolischen Ansätze zur Informationsverarbeitung, welche z.B. von Smolensky [Smo88] mit großen Hoffnungen vertreten werden, aber auch die Idee selbstorganisierender Systeme, wie sie beispielsweise von Maturana & Varela [MV87] vertreten wird. Diese beiden Ansätzen ist in den Abschnitten 8.4 bzw. 8.5 eine eingehendere Betrachtung gewidmet.
Eine weitere neuerdings populär gewordene Idee beruht darauf, Systeme zu entwickeln, die in einer konkreten ('Lebens'-) Umgebung eingebettet werden und dort zunächst einfache Aufgaben bewältigen sollen, z.B. mobile Roboter.[16] Die Fähigkeit zur Lösung solch elementarer Aufgaben, wie die sichere und gegebenfalls zielgerichtete Fortbewegung wird als Voraussetzung für intelligentes Verhalten auf einer höheren Ebene angesehen.[17] Somit repräsentiert dies einen 'bottom-up' Ansatz zur künstlichen Intelligenz, der an die (phylogenetische) Entwicklung biologischer Intelligenz angelehnt ist. Dies steht im Gegensatz zur traditionellen künstlichen Intelligenz die in diesem Sinne einen 'top-down' Ansatz verfolgt; also die abstrakten Intelligenzleistungen zuerst simulieren will, um dann - je nach Bedarf - eventuell schrittweise zu elementareren Leistungen zu kommen.
Im Bereich des maschinellen Lernens, insbesondere Lernen, das auch induktives Schließen beinhaltet, wird derzeit das sogenannte Inductive Logic Programming propagiert.[18] Dieser Ansatz wurde zumindest anfänglich mit großen Hoffnungen beladen und als ein mögliches allgemeines Verfahren zum maschinellen induktiven Lernen betrachtet. Mittlerweile sind die Hoffnungen deutlich gedämpft, wie aus mündlichen Äußerungen Muggletons bekannt ist. Dies bestätigt die Schlußfolgerungen aus den mathematischen Analysen des induktiven maschinellen Lernens in Abschnitt 6.3.

Somit besteht bei der Vorgehensweise der folgende methodologische Zirkel:

> Es werden als schwierig angesehene Aufgaben oder Aufgabengruppen (eventuell unvollständig) definiert. Der Modellentwickler prüft daran eine häufig introspektiv gewonnene Hypothese für die Prinzipien zur Lösung der Aufgabe. Dadurch sind die Prinzipien aber genau auf die gestellten Aufgaben ausgerichtet. Ein allgemeiner Einsatz der Prinzipien ist nicht erfolgreich, weil

[16] Solche Ansätze findet man auch unter dem Stichwort *autonomous systems*.
[17] Siehe für einen neueren Überblick Brooks [Bro91].
[18] Siehe hierzu z.B. Arbeiten von Muggleton [MB88, MF90].

> der Ansatz zu einfach ist (zu geringe Kolmogoroffkomplexität). Somit geben die 'Prinzipien' des jeweils neuen Ansatzes nur wieder, was bei dem Entwurf der Aufgabe bzw. bei dessen menschlicher Lösung vorgegeben wurde. Teilweise wird der Wissenschaftler auch schon an das 'Lösungsverfahren' bzw. die 'Prinzipien' denken, bevor er die Testaufgabe konzipiert.

Mit diesen Problemen wird sich auch Pylyshyn konfrontiert sehen, wenn man beginnt die elementaren Funkionen seiner funktionalen Architektur in empirischen Untersuchungen zu isolieren. Der jeweilige Wissenschaftler wird schon beim Versuchsaufbau seine Vorstellungen von der funktionalen Architektur einfliessen lassen müssen - sonst hätte er überhaupt keinen Anhaltspunkt für die zu variierenden Größen im empirischen Experiment. Dies jedoch impliziert, daß der Erkenntnisgewinn aus den jeweiligen empirischen Untersuchungen äußerst beschränkt ist.

Stimuliert durch die eher enttäuschenden Ergebnisse der bisherigen Forschungsbemühungen in der künstlichen Intelligenz sieht Minsky als einen wesentlichen Aspekt von Intelligenz, daß es sich dabei nicht um ein einheitliches System handelt, sondern um eine Vielzahl von unterschiedlichen Funktionseinheiten, die interagieren.[19]

[19] Siehe hierzu M. Minskys *The Society of Mind* [Min86]. Dieser Gedanke taucht übrigens auch schon bei F. Nietzsche [Nie86] auf, der darauf hinweist, daß die Willensbildung kein einheitlicher Prozeß ist, sondern dabei eine Vielzahl von unterschiedlichen interagierenden Prozessen beteiligt sind. Diese Ideen sind nicht mit dem Konnektionismus zu verwechseln, der seine Stärke in der Vielzahl von *gleichartigen* Funktionseinheiten sieht.

6.3 Über eine formale Theorie des Lernens

In diesem Abschnitt sollen die im vorhergehenden Abschnitt erläuterten allgemeinen methodologischen Probleme in der KI und der Kognitionswissenschaft am Beispiel des Lernens erörtert werden. Dabei wird ausgehend von einer fiktiv vorhandenen allgemeinen Theorie des Lernens erläutert, inwiefern sich ihr Wert umso mehr verringert, je größer der Anwendungsbereich der Theorie gewählt wird.

6.3.1 Formaler Rahmen

Zunächst wird die Schreibweise erläutert und die Definitionen angegeben, die bei den folgenden Untersuchungen benötigt werden.

Es wird eine endliche oder unendliche Objektmenge X betrachtet. Jedes Objekt in X gehört zu genau einer von zwei disjunkten Klassen. Jede Teilmenge von X wird als eine Hypothese bezeichnet. Damit gibt es $2^{|X|}$ extensionsverschiedene Hypothesen. Eine Hypothesenmenge HM ist eine Teilmenge aller Hypothesen über X. Die Elemente einer bestimmten Hypothese H_z, der sogenannten *Zielhypothese* werden *positiv* und die übrigen Elemente in X *negativ* klassifiziert. Als Erfahrungsdaten werden dem Lernenden mit einer Kennzeichnung versehene Objekte aus X präsentiert. Die Kennzeichnung zeigt an, ob das Objekt positiv oder negativ zu klassifizieren ist. Es wird angenommen, daß die Zielhypothese in der betrachteten Hypothesenmenge enthalten ist. Die Lernaufgabe sei für jedes Objekt in X die zugehörige Klasse richtig zu bestimmen. Damit besteht die Lernaufgabe darin, für jedes Objekt in X zu bestimmen, ob es ein Element der anfänglich unbekannten Zielhypothese ist.
Für jede formale Lerntheorie L gibt es dabei in folgendem Sinn genau eine Hypothesenmenge $HM_L \subseteq 2^X$.

a) Für jede Hypothese H aus HM_L gibt es entsprechende Erfahrungsdaten, so daß L lernen wird, die Objekte aus X genau gemäß H zu klassifizieren.

b) eine Lerntheorie L kann niemals eine Hypothese bestimmen, die nicht in der zugrundeliegenden Hypothesenmenge HM_L enthalten ist - gleich welche Erfahrungsdaten präsentiert werden.

Bei relevanten Anwendungen der Lerntheorie werden die jeweiligen Objektmengen sehr groß sein. Angenommen, die Objekte des Lernbereiches werden durch nur 30 unterschiedliche, einstellige Prädikate beschrieben. Dann lassen sich rein syntaktisch bereits $2^{30} \approx 1\,000\,000\,000$ Objektbeschreibungen unterscheiden.

Valiant führte 1984 [Val84b, Val84a] den Begriff des wahrscheinlichen annähernd korrekten Lernens (probably approximately correct learning) ein. Dabei wird eine beliebige, unbekannte aber feste Wahrscheinlichkeitsverteilung D über der Objektmenge X angenommen. Jedes Objekt in X tritt mit einer festen aber unbekannten Wahrscheinlichkeit auf, die durch D bestimmt ist. Während der Sammlung von Erfahrungsdaten für einen

Induktionsschluß werden die Objekte aus X mit einer entsprechenden Klassifikation registriert. Das Ziel eines Induktionsschlusses aus den gesammelten Erfahrungsdaten ist es dann, möglichst alle Objekte aus X richtig zu klassifizieren, die wiederum zufällig nach der gleichen unbekannten Wahrscheinlichkeitsverteilung D auftreten. Dabei ist allerdings - im Gegensatz zu der Situation vor dem Induktionsschluß - keine Klassifizierung der Objekte mehr vorgegeben. Das Ziel dabei ist es, die Wahrscheinlichkeit, daß ein zufällig nach D gewähltes Objekt falsch klassifiziert wird, zu minimieren. Die Wahrscheinlichkeit, daß der Induktionsschluß auf ein zufällig nach D ausgewähltes Objekt nicht zutrifft, wird im folgenden auch die *Fehlklassifikationswahrscheinlichkeit* des Induktionsschlusses genannt. Weiterhin soll mit möglichst großer Wahrscheinlichkeit der Induktionsschluß nur eine kleine Fehlklassifikationswahrscheinlichkeit aufweisen.

Definition 1 *Sei HM eine Hypothesenmenge und $H_z \in HM$ die Zielhypothese. Dann sagen wir, daß eine Lerntheorie L genau dann eine Hypothesenmenge HM* **wahrscheinlich annähernd korrekt lernt***, wenn*

$$(\forall H_z \in HM)(\forall D)(\forall \varepsilon > 0)(\forall \delta > 0)$$

die von L bestimmte Hypothese $H \in HM$ ein zufällig nach D ausgewähltes Objekt $x \in X$ höchstens mit einer Wahrscheinlichkeit von ε fehlklassifiziert. Dieses Ereignis muß dabei mit einer (Konfidenz-) Wahrscheinlichkeit von mindestens $1 - \delta$ auftreten.

Ohne Verlust der Allgemeinheit können noch die folgenden Vereinbarungen getroffen werden. Zunächst wird eine Repräsentation $z(H)$ einer Hypothese H über X definiert.

Definition 2 *Sei jedes Objekt in X durch eine binäre Zahl von 0 bis $|X| - 1$ bezeichnet. Dann heißt die im folgenden beschriebene binäre Zeichenkette $z(H)$ die binäre Darstellung von H. $z(H)$ hat die Länge $|X|$. Zu jeder Position in $z(H)$ korrespondiert ein entsprechend nummeriertes Objekt in X. Eine '1' an der Position für das Objekt $x \in X$ bedeutet $x \in H$, während eine '0' anzeigt, daß $x \notin H$.*

Beispiel: Sei X die Menge $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ und $HM = \{H_1, H_2\}$, wobei
$H_1 = \{1, 4, 8, 9\}$ und $H_2 = \{1, 3, 5, 8, 9\}$ gelte.
Dann ist $z(H_1) =$ '100100011' und $z(H_2) =$ '101010011'.

Damit läßt sich von der Kolmogoroffkomplexität $K(z(H))$ einer jeden Hypothese $H \in HM$ sprechen.

Es folgt die Definition eines Komplexitätsmasses $K_{max}(HM)$ einer Hypothesenmenge HM, welches der Komplexität der komplexesten Hypothese in HM enstpricht:

Definition 3 *Sei K_{max} eine Funktion $2^{2^X} \to \mathbf{N}$ und*

$$K_{max}(HM) = \max_{H \in HM} K(z(H))$$

Dann heißt $K_{max}(HM)$ die Komplexität der Hypothesenmenge HM.

Hypothesenmengen, die komplexe Hypothesen enthalten
Wie weiter oben erwähnt, läßt sich jeder formalen Lerntheorie genau eine Menge von Hypothesen zuordnen. Im folgenden wird die Zahl der Hypothesen betrachtet, die in einer Hypothesenmenge notwendigerweise enthalten sein muß, wenn eine Lerntheorie in der Lage sein soll, auch komplexe Hypothesen zu lernen. Die Zahl der Hypothesen in HM ist interessant, da sie einen Indikator dafür ist, wieviele Erfahrungsdaten benötigt werden, um eine akzeptable Hypothese (geringe Fehlklassifikationswahrscheinlichkeit) aus HM zu bestimmen. Die Erfahrungsdaten dienen dabei dazu, Konkurrenzhypothesen auszuschließen, falls sie inkonsistent mit den Erfahrungsdaten sind.

Theorem 1 *Sei L ein Lernalgorithmus, der binär für eine universelle Turingmaschine U codiert ist. Sei HM_L die Hypothesenmenge, die L zugrunde liegt.*
Für $K_{max}(C) > |L| + \text{const}$ gibt es mindestens

$$2^{K_{max}(HM_L)-|L|-\text{const}}$$

Hypothesen in HM_L, wobei const *eine kleine konstante natürliche Zahl ist, die nur von der konkret gewählten universellen Turingmaschine U abhängt.*

Beweis siehe Hoffmann [Hof90a].

6.3.2 Universelle Lerntheorien

Unter einer *universellen Lerntheorie* wird im folgenden eine formale Theorie verstanden, die es erlaubt, in beliebigen Anwendungsgebieten aus gegebenen Erfahrungsdaten generalisierende Schlüsse zu ziehen. Es wird auf einige generelle Schwierigkeiten bei einem solchen Unterfangen hingewiesen. Die Schwierigkeiten gelten sowohl für deskriptive Theorien des menschlichen Lernens als auch für präskriptive Lerntheorien. Für präskriptive Lerntheorien, die sich mit induktivem Schließen befassen, gibt es eine Reihe von Vorläufern. Beispielsweise versuchte Carnap in *Logical Foundations of Probability* 1950 [Car50] eine formale Induktionstheorie zu entwickeln, die sich an syntaktischen Merkmalen der betrachteten Sätze orientierte. Goodmans Paradoxon[20] ist wohl ein starker Hinweis darauf, daß eine rein syntaktische Betrachtung keineswegs ausreichen kann, um Induktionsschlüsse zu begründen, wie sich in der langen Diskussion herauskristallisierte.[21] Im folgenden wird zwar nicht Carnaps Ansatz weiterverfolgt, obgleich sich eine universelle *formale* Lerntheorie nur an syntaktischen Merkmalen orientieren kann. Goodmans Paradoxon betrifft *epistemologische* Probleme von Induktionstheorien. Im Gegensatz dazu werden im folgenden rein *kombinatorische* Probleme von sowohl präskriptiven als auch deskriptiven Lern- bzw. Induktionstheorien betrachtet. Zunächst werden noch einige Formalien vereinbart.

[20] Siehe z.B. Goodman [Goo55]. Erstmals wurde das Paradoxon 1946 in [Goo46] Goodman veröffentlicht, also noch vor Carnaps Werk.

[21] Siehe z.B. N. Rescher [Res80]. Neuere Arbeiten zum Induktionsproblem finden sich in Stove [Sto86], Indurkhya [Ind90] oder in Holland [Hol86].

Sei n die Zahl der Objekte in X. Dann gibt es 2^n verschiedene Teilmengen oder Hypothesen über X. Weiterhin gibt es 2^{2^n} verschiedene Hypothesenmengen über X.
Das heißt, um eine beliebige Hypothesenmenge beschreiben zu können, werden mindestens $\log_2 2^{2^n} = 2^n$ Binärzeichen benötigt. Eine allgemeine Lerntheorie muß eine konstante und wird in der Regel eine verhältnismäßig geringe Komplexität (Länge) haben. Dabei soll sie auf beliebige Bereiche anwendbar sein.

Für die folgenden asymptotischen Betrachtungen wird weiter von den folgenden Vereinbarungen ausgegangen. Sei X eine abzählbare Menge. Sei $X_N = \{X_i | i \in \mathbf{N}\}$, eine unendliche Menge von Teilmengen von X, wobei gilt: $X_i \subset X_j$, wenn $i < j$. Sei $HM_N = \{HM_i | i \in \mathbf{N}\}$ eine unendliche Menge von Hypothesenmengen wobei gilt: $HM_i \subseteq 2^{X_i}$. Zum Beispiel könnte HM_N die Menge aller Mengen von aussagenlogischen Formeln mit höchstens 17 Disjunktionstermen in disjunktiver Normalform über 1, 2, 3, ... Aussagenvariablen sein.

Das notwendige asymptotische Wachstum von Hypothesenmengen
Im allgemeinen kann man wohl davon ausgehen, daß die zu lernenden Hypothesen bzw. Klassifikationsregeln bei zunehmendem Umfang des Anwendungsbereiches immer komplizierter werden. Die Klassifikationsregeln müssen in den umfangreicheren Bereichen mehr Randbedingungen explizit enthalten, unter denen sie gelten.
Der Zusammenhang aus Theorem 1 läßt sich für asymptotische Betrachtungen wie folgt formulieren:

Theorem 2 *Sei HM_N eine Menge von Hypothesenmengen über X_N. Wenn $K_{max}(HM_i)$ eine in i streng wachsende Funktion ist, dann gilt für alle $i > i_0$ für ein geeignetes i_0: Die Zahl der Hypothesen in HM_i ist mindestens*

$$c(2^{K_{max}(HM_i)})$$

für ein geeignetes $c > 0$.

Beweis siehe Hoffmann [Hof91a].

Obiges Theorem zeigt das Verhältnis von Hypothesenkomplexität zu Hypothesenzahl in einer Hypothesenmenge bei sehr großer Komplexität. das Verhältnis gilt ungeachtet des betrachteten Lernverfahrens. Es zeigt, daß die Zahl der Hypothesen exponentiell mit der Komplexität der kompliziertesten Hypothese ansteigt. Daraus resultiert das folgende Dilemma, bei dem Versuch eine allgemeine formale Lerntheorie aufzustellen, die auch in der Lage ist, auf komplexe Hypothesen zu schließen: Entweder muß diese Lerntheorie sehr lang sein, oder aber sie muß mit einer sehr großen Menge von Erfahrungsdaten versorgt werden, um eine entsprechend komplexe Hypothese zu gewinnen. Anders ausgedrückt bedeutet dies, wenn eine kurz zu beschreibende formale Lerntheorie in der Lage ist, komplexe Hypothesen zu bilden, dann muß die zugrunde liegende Hypothesenmenge auch noch eine Vielzahl anderer, ebenso komplexer Hypothesen enthalten. Dies impliziert,

daß eine entsprechend große Menge an Erfahrungsdaten erforderlich ist, um die vielen Konkurrenzhypothesen der Zielhypothese ausschließen zu können.
Theorem 1 und 2 spiegeln einen fundamentalen Sachverhalt wider. Es zeigt den Zusammenhang auf, zwischen dem Informationsgehalt einer Lerntheorie, der Information in den Erfahrungsdaten und der Information, die in der resultierenden Hypothese enthalten ist. Im folgenden Abschnitt wird gezeigt, was dies für das induktive Schließen aus Beispielen im Rahmen des wahrscheinlich annähernd korrekten Lernens bedeutet.

Induktion aus Beispielen
Es lassen sich verschiedene Arten unterscheiden, wie Erfahrungsdaten für Induktionsschlüsse gesammelt werden. Im folgenden werden die Konsequenzen des obigen Sachverhaltes für das induktive Schließen im Rahmen des *wahrscheinlich annähernd korrekten Lernens* untersucht. Dafür kann die folgende Untergrenze für die Zahl von erforderlichen Beispielen angegeben werden.

Theorem 3 *Sei HM_N eine Menge von Hypothesenmengen über X_N. Sei ferner $K_{max}(HM_i)$ streng wachsend in i. Dann wird für ein* wahrscheinlich annähernd korrektes Lernen *mit einer Fehlklassifikationswahrscheinlichkeit von $1-\varepsilon$ und einer Konfidenzwahrscheinlichkeit von $1-\delta$ mittels einer formalen Lerntheorie L für $i > i_0$ für ein geeignetes i_0, mindestens*

$$c(\frac{K_{max}(HM_i)}{\varepsilon \log |X_i|})$$

Beispiele benötigt, für ein geeignetes $c > 0$ und $0 < \delta < \frac{1}{100}$

Für einen Beweis siehe Hoffmann [Hof91a].

Die Zahl der benötigten Beispiele wächst also asymptotisch mindestens linear mit der Komplexität der kompliziertesten Hypothese in der betrachteten Hypothesenmenge. In der Praxis können Menschen im Gegensatz dazu häufig aus sehr viel weniger Beispielen verlässliche induktive Schlüsse ziehen.[22]

6.4 Diskussion und Schlußfolgerungen

Es wurde in den Theoremen in Unterabschnitt 6.3.2 gezeigt, daß das folgende Problem bei der Entwicklung universeller Lerntheorien vorhanden ist. Je größer die Komplexität der zu bestimmenden Hypothese ist, desto kleiner ist der Einfluß der spezifischen Lerntheorie für die Hypothesenbestimmung. Umgekehrt heißt das, daß mit der Komplexität der zu bestimmenden Hypothese, der erforderliche (algorithmische) Informationsgehalt,

[22] Putnam [Put88] geht ebenfalls davon aus, daß die Beschreibung menschlichen induktiven Schließens sehr umfangreich sein muß. Dies ist gleichbedeutend damit, daß Menschen aus wenigen Beispielen sehr komplexe Aussagen bzw. Aussagensysteme ableiten. Dies aber impliziert gerade das in den Theoremen aufgezeigte Problem mit allgemeinen, universellen Lernalgorithmen.

der aus den Erfahrungsdaten gewonnen werden muß, überproportional ansteigt. So mag eine gegebene Lerntheorie als gut brauchbar für die Erklärung oder Vorhersage von Lernvorgängen in stark eingeschränkten Bereichen sein, sie wird jedoch zunehmend ihren Wert verlieren, wenn die zu lernenden Hypothesen in ihrer Komplexität ansteigen. Für kleine Aufgaben ist das Lernverfahren in der Regel vom Entwickler gerade so entworfen, daß es funktioniert. Es steckt sozusagen ein großer Teil der erforderlichen algorithmischen Information des Lernergebnisses bereits im Algorithmus. Wenn der Algorithmus dann für komplexere Aufgaben eingesetzt werden soll, muß überproportional viel Information aus den Daten gewonnen werden, da der 'Anfangsvorteil' für die Testaufgabe verloren geht. Dies führt dann zu erheblich schlechteren Lernleistungen, als man aufgrund der Experimente in den abgegrenzten Bereichen erwartet hatte. Das Lernverfahren kann also für eine Miniwelt erklären, wie aus wenigen Beispielen richtig generalisiert wird, versagt jedoch bei der Erklärung von Lernprozessen im Alltag. Die Komplexität der unbekannten Zielhypothesen wird regelmäßig ansteigen, wenn die vorhandenen Phänomene und die verwendete Beschreibungssprache der fraglichen Situationen umfangreicher wird. Dann erfordert die Auszeichnung einer bestimmten Phänomenklasse gegenüber den übrigen Phänomenen gewöhnlich eine ausführlichere Beschreibung, z. B. die Angabe von mehr charakteristischen Eigenschaften.
Auch die folgende Beobachtung läßt sich im Rahmen dieses geschilderten Sachverhaltes interpretieren:
Lernprozesse bei Versuchen in Skinnerboxen lassen sich verhältnismäßig gut erklären und vorhersagen, während solche Theorien keine sinnvollen Voraussagen mehr liefern, wenn die fraglichen Lern- bzw. Anpassungsprozesse in komplizierterer Umgebung stattfinden. Dennett [Den71, Den87] fordert aufgrund dieser Beobachtung, eine intentionale Beschreibungsebene für die Erklärung von Verhalten bei komplexen Systemen. Allein durch die Wahl der Beschreibungsebene jedoch ist nach dem *Invariance Theorem* der algorithmischen Informationstheorie noch nichts gewonnen. Der Vorteil der intentionalen Beschreibungsebene läßt sich wohl eher darin sehen, daß dadurch die Nutzung einer sehr umfangreichen Theorie ermöglicht wird, die jedoch nicht in ihrem gesamten Umfang expliziert wird. Der größte Teil der Theorie wird im menschlichen Hintergrundwissen bleiben, das für die Interpretation des dagegen nur kleinen explizierten Teils benötigt wird.
Die intentionale Ebene wirkt dadurch besonders geeignet, da in der menschlichen intersubjektiven Verständigung tatsächlich das Hintegrundwissen nicht expliziert werden muß. Dadurch mag die Beschreibungsebene der intentionalen Zustände fälschlicherweise als Lösungsmöglichkeit des Problems erscheinen. Denn für eine operationale Theorie des menschlichen Denkens oder die Entwicklung eines entsprechenden KI-Systems, muß auch das Hintergrundwissen expliziert werden.

Teil III

Philosophische Probleme und Komplexität

Im dritten Teil der Arbeit soll von den technischen Aspekten der künstlichen Intelligenz abgesehen werden. Stattdessen sollen verschiedene philosophische Fragestellungen unter der Prämisse erörtert werden, daß auch der Mensch in seinem letztlich resultierenden Verhalten eine hochkomplexe Struktur, im Sinne der Kolmogoroffkomplexität, aufweist.

In Kapitel 7 wird die Phänomenologie Martin Heideggers aus *Sein und Zeit* skizziert und die darauf basierende Kritik an der Möglichkeit einer künstlichen Intelligenz, die nach festen, algorithmischen Regeln auf Symbolen arbeiten muß. Diese insbesondere von H. L. Dreyfus vertretene Kritik wird ihrerseits einer Kritik unterzogen, die sich auf die Möglichkeiten einer universellen Turingmaschine bezieht und dabei den Komplexitätsaspekt betont.
Schließlich wird Dreyfus' phänomenologische Kritik als nicht durchschlagend zurückgewiesen, gleichzeitig wird sie aber auf neue Weise als richtungsgebend für die Forschung in der künstlichen Intelligenz interpretiert.

In Kapitel 8 werden mit Heideggers Phänomenologie verwandte Sichtweisen in Bezug auf menschliche Begriffs- und Sprachverwendung, d.h. Wittgensteins Spätphilosophie und Quines Holismusgedanke kurz referiert.
Im Anschluß daran wird in Abschnitt 8.3 analysiert, wie ein komplexes System überhaupt kompakt organisiert sein kann, und inwiefern man etwas, das zu menschlichen Begriffen korrespondieren könnte, dort wiederfindet.
Dabei wird aufgezeigt, daß die Sichtweise Wittgensteins und Quines eine unausweichliche Folge einer hohen Kolmogoroffkomplexitat eines geforderten Systemverhaltens ist.
Abschnitt 8.4 und 8.5 befassen sich mit den oft als Ausweg reklamierten konnektionistischen bzw. selbstorganisierenden Systemen. Auch dort wird wiederum deren Komplexität, gemessen in Kolmogoroffkomplexität, zum Thema gemacht.
Dabei wird deutlich gemacht, daß der konnektionistische Ansatz im Hinblick auf die Handhabung der hohen *problemimmanenten* Komplexität keinen wirklichen Ausweg aus den Problemen des traditionellen symbolischen Ansatzes der KI bietet. Die philosophischen Argumente, die für konnektionistische Systeme sprechen, werden als Scheinargumente entlarvt.
In Abschnitt 8.6 wird schließlich die menschliche Kreativität unter dem Komplexitätsgesichtspunkt analysiert und ihre potentielle Übertragbarkeit auf Maschinen diskutiert.

Kapitel 9 endlich befaßt sich mit der so oft unpräzise gestellten Frage nach den prinzipiellen Grenzen der künstlichen Intelligenz. Hierbei wird in Anbetracht des Komplexitätsaspektes die Frage einer scharfen Präzisierung zugeführt, so daß sie prinzipiell eindeutig beantwortet werden könnte.
Weiterhin wird dort die Komplexität von Systemen angesprochen, die nicht durch Turingmaschinen modelliert werden können.
Kapitel 10 enthält abschließend eine Wiederholung der Hauptpunkte der Arbeit sowie Bemerkungen zu deren Bedeutung für Philosophie, künstliche Intelligenz und Kognitionswissenschaft.

Kapitel 7

Phänomenologie

Die Phänomenologie ist im weiteren Sinne die Wissenschaft von Phänomenen oder Erscheinungen. Da die Gegenstände sich uns jedoch im Bewußtsein offenbaren, heißt Phänomenologie im engeren Sinne die Wissenschaft von den sich im Bewußtsein offenbarenden Phänomenen. Als eine eigene Richtung der Philosophie wurde die Phänomenologie von *Edmund Husserl* begründet. Husserls Ziel war es, eine unanfechtbare Grundlage für alle Wissenschaften zu gewinnen, wozu er die *phänomenologische Methode* entwickelte.[1]

Die phänomenologische Methode teilt sich dabei in zwei Schritte:

1. Die *eidetische Reduktion* sieht zunächst von aller Existenz des *Ich*, der erfassenden Akte und der Gegenstände ab. Vielmehr trachtet die eidetische Reduktion danach, bloß das Wesen (Eidos) der genannten Phänomene in deren voller Konkretion zu erfassen. Dies geschieht durch *eidetische Variation*. Ein Gegenstand wird durch variierende Intentionen angeschaut; z.B. kann ich ein Buch wahrnehmen, verabscheuen, mögen, mich daran erinnern, es mir vorstellen etc.

2. Die *phänomenologische Reduktion* ist der zweite Schritt. Dabei wird auch die Bewußtseinsunabhängigkeit dieser Inhalte eingeklammert. Damit ist ein allgemeines Zweifeln an dem Gegebensein der Gegenstände als methodischer Behelf gemeint. So kann ich beispielsweise an dem Vorhandensein eines Buches zweifeln. Dadurch soll nicht meine Überzeugung von der Existenz des Buches in Frage gestellt werden oder gar negiert werden - wie bei Descartes - sondern lediglich eingeklammert. D.h. die Überzeugung bleibt bestehen, aber es wird eine Distanzierung von den jeweiligen Gegenständen erreicht. Dadurch verlieren sie ihr unmittelbares Gegebensein bei einer natürlichen Einstellung[2], das heißt bei der 'naiven' Betrachtung einer 'realen' Welt. Die konstitutive Rolle des Bewußtseins für die Gegenstände soll dadurch bewußt gemacht werden.

Die Phänomenologie betrachtet ihre Gegenstände immer nur als Korrelate des Bewußtseins. Somit bleibt lediglich das *reine* aber keineswegs leere *Bewußtsein* übrig.

[1] Siehe Husserl [Hus85].

[2] Aus Husserls Ideen zu einer reinen phänomenologischen Philosophie [Hus13] 2. Teil, Bd I.

Nach Husserl gliedert sich das Bewußtsein in Bewußthaben *(Noesis)* und Bewußtes *(Noema)*. Das Bewußte wird erst durch die Noesis als Gegenstand konstituiert. Darum kann das Bewußte (Noema) in unmittelbarer *Wesensschau* oder *Ideation* erfaßt und beschrieben werden. Damit ist die Philosophie als eine rein deskriptive Wesenslehre der immanenten Bewußtseinsgestaltung zu definieren. Da alle Erfahrungsgegenstände durch die ihnen zugrunde liegenden Wesensheiten normiert werden, entspricht jeder empirischen Wissenschaft eine eidetische Wesenswissenschaft oder eine *regionale Ontologie*. Alle *Regionen* (Gegenstandsgebiete) aber gründen ihrerseits im reinen Bewußtsein. Damit ist aber die Philosophie die Wissenschaft vom reinen Bewußtsein und damit auch als *erste* Wissenschaft zu sehen.
Während Husserls Phänomenologie sich vor allem dem Wahrheitsproblem widmete, wandte sich *M. Scheler* der Wertphilosophie zu.[3] An Stelle von Husserls theoretischer Ideation tritt bei Scheler das *Wertfühlen*. Diese muß als nichtverstandesmäßige Erfassung von Werten vorgestellt werden.
Mit *Martin Heidegger*, der Schüler Husserls war, wandelte sich endlich die Phänomenologie radikal zur Existenzphilosophie:

> *Das Wesen des Seins ist nicht überzeitlich ruhendes Bewußtsein, sondern Geschichtlichkeit und Zeit.*

Nach dem Zweiten Weltkrieg wurde die Phänomenologie besonders in Frankreich, Belgien und Nordamerika weiterentwickelt und setzte sich fort in einer *existenziellen* Phänomenologie (z.B. *M. Merleau-Ponty u.a.*).

7.1 Heideggers Philosophie aus *Sein und Zeit*

Heideggers Phänomenologie ist für die vorliegende Arbeit von besonderem Interesse, da sie die Grundlage für heutige philosophische Kritiken an der künstlichen Intelligenz, insbesondere an der *physical symbol system hypothesis* darstellt.[4] Obgleich natürlich nicht die gesamte Systematik Heideggers referiert werden kann, soll versucht werden, die für die Kritik an der KI wichtigen Stellen deutlich zu machen.
Heideggers Konzeption seiner neuen Phänomenologie in seinem umwälzenden Werk *Sein und Zeit* 1927[5] in [Hei27] läßt sich durch die folgenden sechs Stufen beschreiben:

1. Die Grundfrage nach dem Sinn von Sein.
2. Die Klärung der Frage nach dem Sinn von Sein kann nur im Rückgang auf das einzig Seiende, das *Dasein* erfolgen.

[3] In Scheler [Sch13].
[4] Die prominentesten Vertreter der phänomenologischen Kritik an der KI sind Hubert L. Dreyfus und sein Bruder Stuart Dreyfus sowie Terry Winograd und Fernando Flores. Winograd ist selbst ein bedeutender Wissenschaftler der künstlichen Intelligenz, der sich durch seine Kritik von seinem eigenen Forschungsprogramm teilweise distanzierte.
[5] Heideggers Gedanken, die in *Sein und Zeit* niedergeschrieben sind, entwickelten sich seit etwa 1919. Er äußerte sie regelmäßig in seinen Vorlesungen.

3. Das Wesen des Daseins ist das *In-der-Welt-sein.*
4. Das Wesen des In-der-Welt-seins ist die *Sorge.*
5. Das Wesen der Sorge ist die *Zeitlichkeit, wie sie sich in der Sterblichkeit und Endlichkeit - im Tod - manifestiert.*
6. Diese Zeit ist die ursprüngliche Zeit, von der her alle andere Zeit (die Geschichtszeit, die Uhrzeit, die astronomische Zeit etc.) überhaupt erst verstehbar wird.

Die Gliederung deutet an, wie sich Heideggers Idee einer *Hermeneutik der Faktizität* zur vollen Gestalt in *Sein und Zeit* entfaltet hat. Heidegger führt eine völlig neue Terminologie für die Grundbegriffe seiner Philosophie ein. Dabei verwendet er als wesentliche Grundbegriffe *In-der-Welt-sein* und *Sorge.* Die beiden Begriffe, können als das Ergebnis seines Weges zwischen Logik und Leben zur *radikalen Kategorienforschung* angesehen werden. Beide Begriffe drücken seine Abkehr von den vom täglichen Leben losgelösten Wissenschaften aus. Heidegger wendet sich stattdessen hin zu den Fakten der Lebensbewegung des Menschen zwischen Geburt und Tod, so wie sie vor und außerhalb bestimmter Wissenschaften gegeben sind.
Heidegger ist in *Sein und Zeit* der Seinsart des Seienden auf der Spur, in dem sich *Welt* konstituiert. Dabei bricht er mit seinem Lehrer Husserl und der Husserlschen Phänomenologie. Übereinstimmung zwischen beiden besteht zwar darin, daß das Seiende im Sinne dessen, was sie *Welt* nennen, in seiner trasnzendentalen Konstitution nicht aufgeklärt werden kann durch Rückgang auf Seiendes von ebensolcher Art. Heidegger sucht jedoch in seinem neuen Werk nach einer *Fundamentalontologie des Daseins.* Er will aufzeigen, daß die Seinsart des menschlichen Daseins total verschieden ist von der Seinsart alles anderen Seienden. In der totalen Verschiedenheit birgt sie auch gerade die Möglichkeit der transzendentalen Konstitution. Der konkrete Mensch ist als solcher - als Seiendes nie eine *weltlich reale Tatsache*, er ist nicht einfach *vorhanden*, sondern er *existiert.*
Das 'Wundersame', das total Verschiedene, liegt dabei darin, daß die Existenzverfassung des Daseins die transzendentale Konstitution alles Positiven ermöglicht. Das Konstituierende ist etwas und seiend - jedoch von einer anderen Art als der des Positiven. Heidegger fordert also im Gegensatz zu Husserl, daß das Konstituierende in dem *gesamten menschlichen Dasein in einer Welt* gesucht wird. Somit fordert Heidegger also auch ein Abgehen von Husserls *Regionen* oder regionalen Ontologien. Heidegger behauptet gegen Husserls Phänomenologie, daß sich die 'Verfassung' der Weltregionen (d.h. die strukturelle Gliederung und Konstitution der Welt) nicht in einem Bewußtsein vollzieht, sondern im *faktischen, konkreten Existieren.* Husserl wollte den Menschen unter den übrigen Seinsbereichen einordnen, Heidegger hingegen behauptet die totale Verschiedenheit des menschlichen Seins von allem anderen Seienden.
Er geht nicht von einem Bewußtsein aus, das Eindrücke von der Außenwelt empfängt und sich entsprechende Ideen und Vorstellungen macht. Vielmehr fordert er, daß das *Sein*, wie es uns in alltäglichen Lebenssituationen begegnet, und die *Welt* als jeweils praktische Handlungszusammenhänge der Ausgangspunkt sind.

Dies soll an folgendem Beispiel erläutert werden:

> Heute vormittag muß ich zuerst in die Bibliothek und anschließend plane ich noch ein paar Winterschuhe zu kaufen. Ich hoffe, daß ich in der Bibliothek ein geeignetes Buch finde, daß die Bibliothekarin mir schnell weiterhelfen kann, damit ich noch genug Zeit für den Schuheinkauf habe. Hoffentlich finde ich ein passendes Paar, das vor allem genug Halt gibt und bequem ist. Das etwa sind meine Gedanken bei meiner Fahrt in die Stadt, zur Bibliothek. Dann befinde ich mich in der Bibliothek am Register usw.
>
> Später bin ich im Schuhgeschäft und lasse mir von der Verkäuferin die Schuhkollektion zeigen. Ich denke schon an den kommenden Winter und möchte ein Paar Schuhe mit Felleinlage usw.

Jeder könnte die Geschichte, so oder so, weitererzählen. An diesem Beispiel lassen sich grundlegende Einsichten aus *Sein und Zeit* erläutern. Nach Heidegger haben viele philosophische Ansätze diese Einsichten verfehlt, weil sie die *durchschnittliche Alltäglichkeit* übersprungen haben. Sie haben die 'Ferne des Nahen' unterschätzt.
Stattdessen hatte die klassische Analyse der Philosophie etwa Probleme der folgenden Art aufgeworfen: Wie gelangt das isolierte Ich zur Außenwelt ? Wie ist wahre Erkenntnis der Wirklichkeit möglich ? etc.
Heidegger sieht in dieser klassischen Analyse der Bewußtseinsphilosophie mitsamt ihren Fragen eine methodologische Unmöglichkeit.
Denn in der Alltäglichkeit zweifelt niemand an der Existenz der Bibliothek oder der Bibliothekare, und es zweifelt auch niemand daran, daß die Bibliothekarin über ebensolche Wahrnehmungs- und Erkenntnismöglichkeiten verfügt, wie man selbst.
Vielmehr unterstelle man einfach durch sein praktisches Tun, daß die Bibliothekarin existiert und daß man hier nichts bezweifelt. Man zweifelt nicht an der Existenz der Außenwelt aber man behauptet sie auch gar nicht erst. Man fährt einfach nur in die Stadt, geht einfach in die Bibliothek oder in das Schuhgeschäft ohne irgendetwas zu hinterfragen.

Zunächst läßt sich also festhalten, daß wir uns in primären praktischen Lebenssituationen wie selbstverständlich 'in der Welt' bewegen. Und ebenso leben wir mit den anderen, 'mit den Mitmenschen' gemeinsam in diesen Situationen. Diese Grundzüge des menschlichen Lebensvollzugs, das *In-der-Welt-sein* und das *Mit-sein-mit-anderen*, nennt Heidegger *Existenzialen*. Sie bezeichnen keine Eigenschaften einzelner Subjekte, sondern die *Form* von komplexen Lebensvollzügen. Somit sind Feststellungen wie 'Ich bin auf der Welt' völlig verschieden von Feststellungen wie 'Ich habe braune Augen' oder 'ich bin 1,72m groß'. Die beiden letzten Feststellungen betreffen objektive Eigenschaften, dingliche Qualitäten meines Körpers. Die Begriffe für solche Qualitäten nennt Heidegger *Kategorien*. *Nichtmenschlich* Seiendes läßt sich demnach *nur-kategorial* beschreiben. Wenn ich aber sage, daß ich auf der Welt bin, dann betrifft diese Feststellung die *Form meines Lebens* und keine Eigenschaft meiner Person, meines Bewußtseins.
Vor allem aber ist mein In-der-Welt-sein keine Tatsache in meinem Leben oder *in meiner Welt.* Daß ich in der Welt lebe, diese 'ontologische', 'formale' bzw. 'existenziale'

Tatsache, das sehe ich an nichts in der Welt. Vielmehr ist es die Voraussetzung - die transzendentale Konstitutionsbedingung - für *Welt* überhaupt.
Ebenso wie nichts in meinem Gesichtsfeld darauf hindeutet, daß es von einem Auge aus gesehen wird.[6] Das In-der-Welt-sein bedeutet, jeweils (immer schon) in bestimmten bedeutungsvollen praktischen Lebenssituationen zu sein. Wenn ich z.B. jetzt beim Anprobieren im Schuhgeschäft bin, so begegnet mir *innerweltlich Seiendes*, nämlich jeweils ein Schuh 'zum Anprobieren'. In den primären Lebenssituationen handelt es sich nicht um eine theoretische Gegenständlichkeit, die ich bloß theoretisch anschaue. Das wäre das Verhältnis zu etwas *Vorhandenem*. Ich befinde mich aber nicht in einem solch distanzierten Verhältnis zu der mich umgebenden Außenwelt. Sondern ich befinde mich gerade bei den Schuhen, bei der Anprobe. Die Schuhe werden von der Verkäuferin gereicht und ich probiere sie an und prüfe, ob sie auch ausreichend bequem sind. Heidegger nennt solches Seiende *Zeug*, und seine Seinsart *Zuhandenheit*. Die Welt begegnet zunächst stets in solchen praktischen Lebenszusammenhängen. Das *Zeug*, hier Schuhe, Schuhlöffel, der Stuhl auf dem ich sitze usw. - bildet einen *Verweisungszusammenhang*. Die einzelnen Gebrauchsgegenstände verweisen aufeinander, weil sie eine *Bewandtnisganzheit* bilden. Das Schuhgeschäft ist gerade so eingerichtet, daß man sich beraten lassen, Schuhe auswählen, anprobieren und kaufen kann. Die Bibliothek ist so eingerichtet, damit man in ihr studieren und nach Literatur recherchieren kann, sie enthält Kataloge, Arbeitsplätze, es sind sachkundige Bibliothekare da, usw.
Diese Struktur des *um-zu* konstituiert die gesamte existenzielle, praktische Räumlichkeit und auch Zeitlichkeit. Wenn wir uns in der von Menschen bewohnten Welt umblicken, so wird der Gebrauchssinn der Räume und der Zeiten offenkundig. Dort ist der Parkplatz, dort die Straßenbahnhaltestelle und dort hinten die Einkaufsstraße. Die Geschäfte liegen in der Einkaufsstrasse dicht beieinander, damit man die alltäglichen Besorgungen unmittelbar nacheinander erledigen kann.

Somit ist festzuhalten:

1. Das menschliche Dasein, seine praktischen Lebensvollzüge, bilden die Basis zum Verständnis der Welt und allem übrigen Seienden.

2. Das Dasein ist immer schon in der Welt *aufgegangen*, indem es sich in ihr tätig orientiert, um jeweils etwas zu erreichen, etwas zu besorgen, etwas zu vermeiden, usw.

3. Das zunächst begegnende Seiende hat für das Dasein die Seinsart der *Zuhandenheit*. Es ist zu etwas gut und wird dazu benutzt, um dieses oder jenes zu tun.

Die primäre Art und Weise, *in-der-Welt* zu sein, nennt Heidegger daher die *Sorge* - das tätige Umgehen mit etwas. Beispielsweise stehe ich morgens auf und gehe ins Badezimmer, um mir die Zähne zu putzen. Normalerweise würde ich die Zahnbürste nicht erst noch

[6]Dies ist eine Formulierung aus Wittgensteins Tractatus [Wit21] der zu ähnlichen Schlüssen in dieser Hinsicht gelangt - allerdings gelangte er von einem ganz anderen Ausgangspunkt dorthin.

besonders bemerken, sondern dieses *Zeug* einfach *zur Hand* nehmen, *um* mir die Zähne zu putzen. Es bestünde ein einfacher Bewandtniszusammenhang. Wenn nun plötzlich keine Zahnbürste *da* ist, dann erst bemerke ich eigens, daß sie überhaupt *da*, im Sinne von *vorhanden* war. Erst durch die *Störung der Verweisung* wird die Verweisung *ausdrücklich*. Je 'reibungsloser' alles funktioniert, desto fragloser gehe ich *in der Welt* auf. Ja, ich bin von ihr *benommen*. Insofern macht die ursprüngliche, tätige Weltvertrautheit unser Seinsverständnis aus.

Damit hat Heidegger also dargelegt, daß

- es keinen Sinn hat, ein von seiner Welt isoliertes *Ich* oder *Bewußtsein* oder *Subjekt* in Ansatz zu bringen.
- daß nicht ein theoretisches, ein Weltverhältnis der bloßen *Vorhandenheit* für das menschliche Dasein grundlegend ist.

Vielmehr gilt:

- Dasein ist immer schon in einer Welt, *bei der Welt*.
- Der tätige Umgang geht stets z.B. theoretisch betrachteten Weltverhältnissen voraus.

Die *Welt*, *meine* Welt, gliedert sich zunächst räumlich und zeitlich in sinnvolle Verweisungszusammenhänge.
Im alltäglichen Umgang weist das *Zeug*, das zur Besorgung bestimmter Dinge *zuhanden* ist, immer auch schon auf andere hin, die es hergestellt haben oder es ebenfalls gebrauchen können. Beispielsweise ist die Bibliothek mit ihrem Register und den Bibliotheksangestellten gerade so eingerichtet, daß sie nicht nur von mir, sondern auch von anderen benutzt werden kann. Sie ist für die Zwecke die *man* in einer Bibliothek verfolgt, eingerichtet. Diese Öffentlichkeit der alltäglichen Lebenssituation nennt Heidegger auch *die Ausgelegtheit durch das Man*. Das großgeschriebene *Man* bezeichnet dabei das *Wer* der Alltäglichkeit. Das *Man* bügelt gleichsam alles Große und Bedeutende nieder, macht es sich für seine banale Durchschnittlichkeit zurecht.

> *Jeder Vorgang wird geräuschlos niedergehalten. Alles ursprüngliche ist über Nacht als längst bekannt geglättet. Alles Erkämpfte wird handlich. Jedes Geheimnis verliert seine Kraft. Die Sorge der Durchschnittlichkeit enthüllt wieder eine wesenhafte Tendenz des Daseins, die wir die Einebnung aller Seinsmöglichkeiten nennen.*[7]

Geworfenheit nennt Heidegger die Art, wie das *ich* zu seinem eigenen *In-der-Welt-sein* gekommen ist.

[7] Aus *Sein und Zeit* §27.

Die Geworfenheit ist nicht die faktische Geburt, sondern die konstitutive Form jedes menschlichen Lebens. Der von Heidegger gewählte Ausdruck deutet an, daß wir ungefragt und ohne persönliche Zustimmung in die Welt gekommen sind. Diese Geworfenheit, dieses ungefragte In-die-Welt-gekommen-sein, ist die *Form*, die die Faktizität des Daseins in *Sein und Zeit* annimmt.
Heidegger nennt sie auch die (konstitutive) Tatsache, *sein Da sein zu müssen.*
Unser jeweiliges *Da* sein, das können wir im konkreten Lebensvollzug nur so, daß wir selbst handeln. Im großen Maßstab, auf unser ganzes Leben bezogen, so Heidegger, existieren wir. Das Dasein aber hat zweierlei:

- Einerseits ist es durch seine ungefragte Geworfenheit (Faktizität) gekennzeichnet,
- und andererseits hat es auch Entwurf-Charakter (Existenzialität).

Die Geworfenheit meint das *Da-sein müssen*, während die Existenzialität, der Entwurf-Charakter, das *Da-sein können* ausdrückt.
Im Sich-Entwerfen nimmt das Dasein gleichzeitig ein Verhältnis zu sich selbst ein: indem es *sein Sein zu sein hat*, verhält es sich selbst zu seinem Sein: Es geht ihm um sein eigenes Sein.
Diese beiden Grundzüge: Faktizität und Existenzialität, faßt Heidegger mit der Formulierung zusammen, Dasein sei *geworfener Entwurf.*[8]
Ein drittes Strukturmoment tritt als die *Verfallenheit* in *Sein und Zeit* auf. Verfallenheit meint, daß das Dasein als geworfener Entwurf sich entwerfend gleichzeitig bereits an die Alltäglichkeit des *Man* gebunden - *verfallen* - ist. Durch das *Man*, durch die Öffentlichkeit der Verweisungszusammenhänge ist es bereits *entfremdet.* Beispielsweise zeigt sich diese Entfremdung bereits in der Schuhkollektion, die das Schuhgeschäft anbietet. Die Schuhe sind so entworfen und auch ihre Größen so ausgewählt, daß auch andere etwas Passendes finden. Die Schuhverkäuferin wird mich bei der Anprobe vielleicht darauf hinweisen, daß ein bestimmter Schuh besonders häufig gekauft wird. Dadurch wird auch meine Entscheidung so oder so beeinflußt werden.

Heideggers *Dasein* läßt sich also in die drei gleichursprünglichen Momente

- Existenzialität,
- Faktizität,
- Verfallenheit

aufgliedern.
Heidegger fragt also *nach der ursprünglichen Einheit des Strukturganzen von Dasein.*

[8] Aus *Sein und Zeit* §31.

> *Die Geworfenheit aber ist die Seinsart eines Seienden, das je seine Möglichkeiten selbst ist, so zwar, daß es sich in und aus ihnen versteht (auf sie sich entwirft). ... Das Selbst aber ist zunächst und zumeist uneigentlich, das Man-selbst. Das In-der-Welt-sein ist immer schon verfallen. Die durchschnittliche Alltäglichkeit des Daseins kann demnach bestimmt werden als das verfallend-erschlossene, geworfen-entwerfende In-der-Welt-sein, dem es in seinem Sein bei der Welt und im Mitsein mit anderen um das eigenste Seinkönnen selbst geht.*[9]

Die *Befindlichkeit*, die Gestimmtheit ist eine Form der Faktizität. Heidegger zählt die Befindlichkeit zu den transzendentalen Möglichkeitsbedingungen des In-der-Welt-seins. Die Welt ist mir immer in einer bestimmten Stimmung *erschlossen*, das heißt zugänglich und verstehbar. Ich kann fröhlich, traurig, müde, munter usw. sein, während ich mein In-der-Welt-sein erlebe und Besorgungen mache. Die *Furcht* ist ebenfalls eine solche Stimmung, ein Modus der Befindlichkeit. Das Wovor der Furcht ist dabei immer ein *innerweltlich Begegnendes*.[10] Man kann sich beispielsweise vor Arbeitslosigkeit fürchten oder davor, daß man nicht die richtigen Schuhe im Schuhgeschäft bekommt.
Die *Angst* hingegen, hat kein innerweltlich Seiendes als *Wovor* der Angst. Das Wovor der Angst ist vielmehr das In-der-Welt-sein als solches.

> *Worum sich die Angst ängstet, ist das In-der-Welt-sein selbst. In der Angst selbst versinkt das umweltlich Zuhandene, überhaupt das innerweltlich Seinende. Die 'Welt' vermag nichts mehr zu bieten, ebensowenig das Mitdasein anderer. Die Angst benimmt so dem Dasein die Möglichkeit, verfallend sich aus der 'Welt' und der öffentlichen Ausgelegtheit zu verstehen. Sie wirft das Dasein auf das zurück, worum es sich ängstet, sein eigentliches In-der-Welt-sein-können.*

Heideggers Zeitlichkeitsanalyse wird durch die *Gleichursprünglichkeit* von Aspekten der Lebensphänomene deutlich.
Einfache Beispiele hierfür sind ganze Sätze, in denen sich kein Wort entfernen läßt, ohne den Satzsinn zu zerstören: 'Dieses Auto fährt schnell.' In diesem Satz bildet sich der Sinn *gleichursprünglich* aus allen vier Worten. Dies meint Heidegger entsprechend angewendet auf ganze Lebenssituationen. Man stelle sich beispielsweise ein Foto vor, auf dem zwei Staffelläufer bei der Übergabe der Staffel zu sehen sind. Der eine Läufer reicht dem anderen die Staffel, der andere streckt seinen Arm aus, um die Staffel zu übernehmen.
Ich kann diese - auf dem Foto erstarrte - Lebensbewegung nur verstehen, wenn ich sie *nach hinten* und *nach vorne* fortzusetzen vermag. Die obige verständliche Beschreibung des Fotos - ohne es vorzuzeigen - setzt schon diese minimale Ganzheit einer jeder Lebenssituation voraus.

[9] Aus *Sein und Zeit* §39.
[10] Aus *Sein und Zeit* §30.

Husserl hatte solche Ganzheit anhand von Beispielen aus der Musik erläutert. Das Hören einer Melodie ist nicht auf Hören jeweils eines einzelnen Tones zu reduzieren. Um eine Melodie zu vernehmen, muß ich (mindestens) jeweils noch den gerade verklungenen Ton *im Ohr* haben und, indem ich den jetzt erklingenden Ton höre, muß ich schon den (möglichen) nächsten Ton *in Gedanken vorwegnehmen*. Das Vorwegnehmen nennt Husserl die *Protention* des inneren Zeitbewußtseins und das Nachklingen die *Retention*.
Mit seinen Analysen zur Gleichursprünglichkeit in der existenzialen Konstitution des In-der-Welt-seins weitet Heidegger solche, zunächst auf einzelne Wahrnehmungssituationen bezogenen phänomenologischen Untersuchungen auf das Ganze des menschlichen Selbst- und Weltverhältnisses aus.
Das Zeitliche geht somit gleichursprünglich aus dem Zukünftigen, dem Gegenwärtigen und dem Vergangenen hervor.

7.2 Die phänomenologische Kritik an der KI

Die Heideggersche Phänomenologie und die KI wurden durch Dreyfus' Veröffentlichung *Alchemie und KI* [Dre65] bereits in den 60er Jahren miteinander in Verbindung gebracht. Doch erst in den 80er Jahren fand diese Verbindung eine breitere Fachöffentlichkeit. Das größere Interesse in der Fachwelt ist unter anderem durch Winograds Abkehr vom traditionellen Forschungsprogramm der KI und seinem Buch *Understanding Computers and Cognition* [WF86] zu erklären. Im wesentlichen wird von Dreyfus [Dre72, DD87, DD88], Winograd & Flores [WF86] und anderen die folgende Kritik an dem klassischen, dem symbolischen Ansatz der künstlichen Intelligenz angebracht:
Es wird argumentiert, daß für symbolmanipulierende Systeme Informationen[11] bloß *vorhanden* sind. Maschinen müssen die Vorhandenheit einer fixierten 'Welt' repräsentieren und können nur die repräsentierenden Symbole nach starren, festen Regeln manipulieren. Im Gegensatz dazu macht aber im menschlichen Denken und Handeln die Zuhandenheit den weitaus größten und wichtigsten Teil aus.[12] Eine externe Welt wird gar nicht erst bewußt gemacht. Menschliches Wissen ist zu einem großen Teil nur implizit im menschlichen Handeln, in dessen spezifischer Struktur enthalten - und nicht symbolisch repräsentiert oder repräsentierbar, so Dreyfus.
Diese Beobachtung reduziert Dreyfus zunächst auf die Unterscheidung zwischen *knowing-how* und *knowing-that*. Mit *knowing-how* ist das Handlungswissen gemeint, das sich nur in der tatsächlich durchgeführten Handlung *zeigt*. Dieses Wissen ist - da es sich nicht durch Symbole repräsentieren läßt - einem Formalismus auch nicht mitteilbar - so Dreyfus.
Demgegenüber steht das Wissen von der Art des *knowing-that* welches durch Symbole repräsentierbares Wissen meint.

[11] 'Information' ist hier als vorwissenschaftlicher Begriff gemeint. Man sollte vielleicht besser von 'Daten' sprechen.

[12] Beispielsweise ist das (motorische) Wissen um die Bedienung eines PKW bei dem geübten Fahrer schlicht *zuhanden*. Das heißt, der Fahrer wird sich der *Vorhandenheit* gar nicht mehr bewußt. Vielmehr zeigt sich die Zuhandenheit des Wissens einfach im Vollzug des tatsächlichen, konkreten Wagenlenkens.

Wissen über Beziehungen von Objekten in einer *vorhandenen* Welt. Solche Welten treten jedoch im wesentlichen nur im Falle von Heideggers Verweisungsstörungen auf und sind dementsprechend kontextabhängig. Je nach Situation wird eine andere Ontologie *vorhanden*.

Dreyfus behauptet also zunächst die Wesensverschiedenheit von *knowing-how* und *knowing-that* und die entsprechende Dichotomie die sich damit über den verschiendenen Arten von Wissen aufspannt.[13]
Dies mag überraschen, da Dreyfus Heideggers *Sein und Zeit* im wesentlichen mit Wittgensteins Spätphilosophie gleichsetzt. Dabei weist er darauf hin, daß der Ansatz der traditionellen KI eigentlich Wittgensteins Frühphilosophie aus dem Tractatus entspricht. Wittgenstein macht jedoch bereits im Tractatus auf die Unterscheidung zwischen sprachlich Beschreibbarem und nicht Beschreibbarem aufmerksam:

> Es gibt allerdings Unaussprechliches. Dies *zeigt* sich, es ist das Mystische.[14]

Weiterhin behauptet Dreyfus für die nicht symbolisch repräsentierbare Wissensart des *knowing-how*, daß sie eines nicht durch algorithmische Regeln beschreibbaren Inhaltes ist. Das Verhalten ist nur im philosophischen Sinne regelgeleitet - folgt also nur Regeln mit nicht näher bestimmbaren Ausnahmen.
Daraus schließt Dreyfus letztendlich, daß menschliches Intelligenzverhalten nicht algorithmisierbar ist - da es auch wesentlich Wissen von der nicht algorithmisch beschreibbaren Art des *knowing-how* involviert.
Mithin eine künstliche Intelligenz nie Leistungen hervorbringen wird, die mit menschlichen Intelligenzleistungen vergleichbar sein könnten.

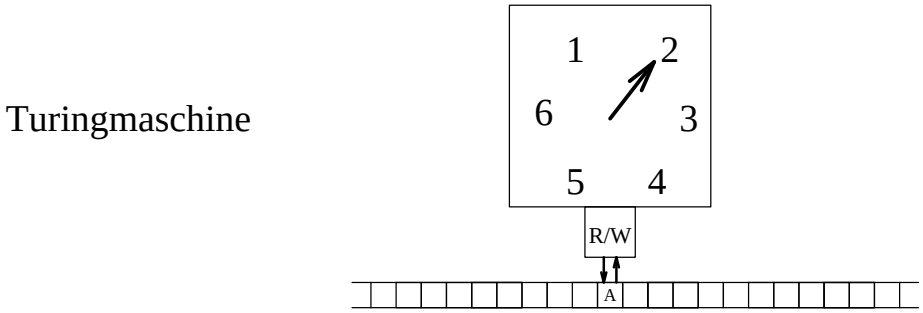
7.3 Trifft Dreyfus' Kritik die klassische KI ?

Dreyfus' phänomenologische Kritik an der *phyiscal symbol system hypothesis* will aufzeigen, daß bestimmte phänomenologische Begriffe kein Pendant bei Algorithmen - bei einer künstlichen Intelligenz haben. Daraus folgert er die Unmöglichkeit einer symbolischen KI. Hierbei sei zunächst angemerkt, daß man von einer Korrespondenz zwischen den phänomenologischen Termini Heideggers und Eigenschaften von Algorithmen oder Computern bestenfalls metaphorisch sprechen kann, da Algorithmen das Bewußtsein fehlt, welches der wesentliche Gegenstand der Philosophie Heideggers ist. Im folgenden wird gezeigt, daß die Unterscheidung zwischen *knowing-how* und *knowing-that* <u>nicht</u> die von Dreyfus hervorgehobenen Konsequenzen impliziert. Diese lauten, daß Dreyfus' Unterscheidung

1. eine Dichotomie zwischen menschlichen Wissensarten etabliert.
2. nachweist, daß regelbefolgende Maschinen an dem *knowing-how* scheitern müssen.

Zu diesem Zweck soll zunächst eine Turingmaschine näher betrachtet werden. Siehe Abbil-

[13] Hucklenbroich [Huc89] zeigt auf, daß sich hier keine strenge Dichotomie aufrecht erhalten läßt.
[14] Wittgenstein [Wit21] Absatz 6.522.



Turingtabelle

Zustand	x	Zeichen	→	Zustand	x	Zeichen	x	Kopfbew.
1		A	→	2		A		L
1		B	→	3		A		R
1		C	→	6		D		R
.		.	.	.		.		.
.		.	.	.		.		.
.		.	.	.		.		.
.		.	.	.		.		.
.		.	.	.		.		.
6		D	→	3		B		L

Abbildung 7.1: Das Schema einer Turingmaschine und die zugehörige Turingtabelle.

dung 7.1. Die Funktionsweise der Turingmaschine ist dabei durch die zugehörige Turingtabelle definiert. Das heißt, die jeweils nächste Aktion der Turingmaschine ist abhängig von ihrem jeweiligen internen Zustand und dem Zeichen, das sich auf dem Band unter dem Schreib-/Lesekopf befindet. Abhängig von diesen beiden Komponenten bestimmt eine entsprechende *Regel*, die in der Turingtabelle angegeben ist, die jeweils nächste Aktion. Dazu gehört, welches Zeichen auf das Band geschrieben wird, welchen internen Zustand die Turingmaschine als Nächstes einnimmt und wie sich der Schreib-/Lesekopf auf dem Band bewegt.
Wenn man die Turingmaschine als 'Black box' in ihren Aktionen beobachtet, so läßt sich feststellen, daß sie die einzelnen Zeichen auf dem Band einfach manipuliert. Ihre Aktionen geschehen einfach - ohne explizite Reflexion oder Regelanwendung. Ihr 'Wissen' *zeigt* sich in ihrem Verhalten und ist implizit in der Struktur ihres faktischen Verhaltens enthalten. Diese Verhaltensstruktur entspricht dabei genau dem phänomenologischen Argument, das Dreyfus nur auf Menschen bezogen wissen will. Die Turingmaschine vollzieht ihre Verarbeitungsprozesse *einfach*, ohne einer ausdrücklichen Regelanwendung zu bedürfen. Dies geschieht gleichermaßen, wie Menschen einfach Radfahren oder die Gangschaltung eines PKW bedienen; ohne erst über irgendwelche Regeln zur Hand- oder Fußbewegung zu räsonieren.
Weiterhin läßt sich beobachten, daß sich die jeweils *einfach vollziehenden Aktionen* der Turingmaschine sich präzise in der jeweiligen Turingtabelle *explizieren* lassen !
Jedoch genügt in der Tat nicht die bloße Niederschrift der Regeln, denn diese müssen natürlich geeignet *interpretiert* werden. Im Falle des rein theoretischen Modells der Tu-

ringmaschine übernimmt der Mensch die Interpretation der Turingmaschinentabelle. Bei physikalischen Realisierungen einer Turingmaschine[15] übernehmen entsprechend entworfene physikalische Prozesse die angemessene Interpretation der Turingmaschinentabelle. Hierbei ist hervorzuheben, daß man sich bei dem 'schlichten Vollzug' der Aktionen der Turingmaschine ein beliebig komplexes Verhalten vorstellen kann. Mithin kann ein beliebig komplexes 'implizites Wissen' durch das schlichte Verhalten der Turingmaschine realisiert werden, wenn die Turingmaschinentabelle nur hinreichend lang ist und ein komplexes Verhalten vorschreibt. In diesem Sinn wäre das Verhalten durch ein *knowing-how* bestimmt - und nicht durch ein *knowing-that* !

7.3.1 Zur Interpretation einer Turingmaschinentabelle

Die Interpretation der Turingmaschinentabelle kann sogar ihrerseits durch eine besondere Turingmaschine geschehen. Die *universellen Turingmaschinen* sind in der Lage eine beliebige Turingmaschinentabelle zu interpretieren. Und zwar in dem Sinn, daß die durch die Tabelle beschriebene Turingmaschine vollständig *simuliert* wird.[16] Eine universelle Turingmaschine kann jede beliebige Turingmaschine simulieren - sogar eine andere universelle Turingmaschine oder sich selbst.
Somit kann der Entdeckung der Existenz von *universellen Turingmaschinen* eine überragende Bedeutung im Hinblick auf Dreyfus' Auslegung von Heideggers Phänomenologie in diesem Zusammenhang beigemessen werden !

> Denn die universelle Turingmaschine repräsentiert in diesem Sinne ein 'universelles implizites Wissen', ein *knowing-how*, das es erlaubt, beliebiges anderes 'implizites Wissen' als explizites Wissen, als *knowing-that*, darzustellen !

Generell läßt sich allerdings festhalten, daß die Repräsentation einer Turingmaschinentabelle auf dem Band einer universellen Turingmaschine nur durch Zeichen geschehen kann, die auf *abstrakte* Entitäten verweisen. Die Zeichen repräsentieren also keine ontologischen Entitäten einer physischen Umwelt, wie es bei der Argumentation von Dreyfus dem als *explizit* bezeichneten menschlichen Wissen (dem *knowing-that*) zukommt.

Eine andere Frage, die sich hier allerdings anschließt, und die von Dreyfus sicherlich als problematisch betrachtet wird, ist die folgende: Auf welcher Beschreibungsebene sollen oder müssen bei Computern die entsprechenden Prozesse beschrieben werden ? Und inwiefern gelingt es Menschen ihre eigenen kognitiven Prozesse auf der jeweiligen Ebene zu beschreiben ? Darauf wird noch in Abschnitt 7.4 näher eingegangen.

[15] Eine physikalische Realisierung des theoretischen Modells einer Turingmaschine ist genau genommen natürlich nicht möglich, da ein *unendliches* Schreib-/Leseband physikalisch nicht realisiert werden kann.
[16] Vergleiche Abschnitt 3.3.

7.3.2 Warum wirkt die phänomenologische Kritik so überzeugend ?

Im folgenden werden zwei Gründe für die Überzeugungskraft der phänomenologischen Kritik erörtert. Diese Gründe zeigen auf, inwiefern die Kritik neue nützliche Sichtweisen auf einige Probleme der KI eröffnet.

- Zunächst sollen intelligente Systeme aus den beiden Perspektiven Benutzer versus Entwickler betrachtet werden.
- Der zweite Aspekt betrifft die Komplexitätsdifferenz zwischen unbewußten und bewußten Prozessen im menschlichen Denken.

Möglicherweise beruht der gesamte Disput um die *physical symbol system hypothesis* einfach auf den beiden unterschiedlichen Sichtweisen: Der Systementwickler auf der einen Seite und der Systembenutzer auf der anderen Seite.
Die Befürworter der Hypothese fragen danach, ob ein physikalisches symbolmanipulierendes System *möglich* ist, das intelligentes Verhalten simuliert. Hingegen fragen die Gegner der physical symbol system hypothesis, ob ein pysikalisches symbolmanipulierendes System *entwickelt* werden kann, das intelligentes Verhalten simuliert: genauer, das Symbole derart manipuliert, wie es die bewußten menschlichen Denkprozesse zu tun scheinen.
Denn die menschlichen bewußten Denkprozesse sind in der Tat wesentlich durch implizites Wissen bestimmt, statt schematisch Symbole zu manipulieren, die explizit bestimmte Wissensfragmente repräsentieren. Dies leitet sich bereits aus der Tatsache ab, daß das menschliche Bewußtsein in seiner Fassungskapazität sehr beschränkt erscheint, während die menschlichen Denkprozesse insgesamt von großer Kolmogoroffkomplexität zu sein scheinen.[17]
Wenn man das menschliche Wissen *operationalisieren* will - und dies tut der Entwickler von intelligenten Systemen im wesentlichen - so ist man daher typischerweise mit der Implizitheit menschlichen Wissens in den unbewußten Denkprozessen konfrontiert.
Die phänomenologische Kritik betont letztlich lediglich den starken Einfluß dieses unbewußten Wissens - und dessen enorme Komplexität - auf die bewußtwerdenden Gedanken und damit auf das menschliche Verhalten.
Daher mag der Schluß überzeugen, daß pysikalische symbolmanipulierende Systeme nicht adäquat sein können, wenn sie immer nur Symbole manipulieren, die etwas repräsentieren, das uns Menschen durch Introspektion[18] bewußt wird. Denn das Bewußtwerdende reicht in der Tat nicht aus, um unsere Denkprozesse vollständig zu beschreiben.[19]

Auf der anderen Seite, der Benutzerseite, sieht die Lage etwas anders aus: Hier stellt sich die Frage, ob es ein System geben kann, das im wesentlichen den Turingtest besteht - das nämlich in einer bestimmten Klasse von Situationen intelligente oder nützliche Reaktionen zeigt. Bei dieser äußerlichen Problembeschreibung ist es schwer einzusehen, warum

[17] Vergleiche auch mit Abschnitt 8.3.
[18] Mit *Introspektion* ist hier eine Selbstbeobachtung der im Bewußtsein stattfindenden kognitiven Prozesse gemeint, mit der Absicht, diese Prozesse erklären zu wollen.
[19] Dies geht auch aus Wittgensteins Regelfolgen hervor. Vergleiche Abschnitt 8.1.

es keine algorithmische Beschreibung für ein nützliches Antwortverhalten geben können sollte.
Die Gegner der *physical symbol system hypothesis* hingegen, haben lediglich mit der phänomenologischen Kritik dargelegt, daß eine direkte Vorgehensweise bei der Entwicklung intelligenter Systeme nicht möglich ist. Versucht man, die Ontologie, die ein Experte in einem bestimmten Problembereich äußert, durch Symbole zu repräsentieren, so ist man damit noch lange nicht am Ziel angelangt - so die phänomenologische Kritik. Genauergenommen behautet die phänomenologische Kritik, daß die verbleibende Strecke zum Ziel *überhaupt nicht* überwunden werden kann. Jedoch bietet sie an dieser Stelle keine plausible Argumentation dafür an, daß der Rest der Strecke nicht doch auf irgendeine Weise algorithmisiert werden kann. Diese Fragestellung wird schlicht unterschlagen.
Der von der phänomenologischen Kritik hervorgehobene Aspekt der wechselnden Ontologie - je nach Situation - ist ebenfalls kein Einwand für einen Benutzer. Der Benutzer wird in der Regel keine Schwierigkeiten haben, sich vorzustellen, daß bestimmte interne Zustände eines Systems Unterschiedliches repräsentieren können - je nachdem, wie das System gegenwärtig eingesetzt wird.[20] Letzteres wird allerdings für den Entwickler zum Problem, wenn er alle möglichen Interpretationen interner Zustände bereits während der Systementwicklung vorhersehen soll. Dies ist jedoch auf die Entwicklerperspektive beschränkt.
Die Bedeutung eines systeminternen Zustands ist nur solange eine feste Bedeutung, solange sie bloß *innerhalb* des Systems definiert werden soll. Dabei bleibt die Rolle des Systems inmitten einer 'Lebenswelt', z.B. einer menschlichen Gemeinschaft, außer Betracht. Hingegen kann die Benutzerinterpretation der Systemzustände - je nach Systemeinsatz - wechseln.

Die Komplexitätsdifferenz zwischen bewußten und unbewußten Denkprozessen wirft das folgende Problem auf: Da die bewußten Denkprozesse - solange sie überhaupt algorithmisch erscheinen - nur von geringer Kolmogoroffkomplexität sein können, müssen die unbewußten Denkprozesse von entsprechend hoher Kolmogoroffkomplexität sein. Sozusagen ist die 'Maschinenebene' (Pylyshyns funktionale Architektur) von weitaus komplexerer Gestalt als eine einfache universelle Turingmaschine oder die Maschinenebene eines realen Computers. Dies läßt die mögliche Korrespondenz der menschlichen unbewußten Denkprozesse und einem Computer schwer erkennen.[21] Der Computer ist ja tatsächlich viel einfacher strukturiert und muß erst durch ein entsprechend komplexes Programm, das dabei nur zu einem kleinen Teil zu den bewußten Denkprozessen korrespondiert, zu einem vergleichbar komplexen System gemacht werden.
Dadurch ist die mögliche Simulation der sehr komplexen unbewußten Denkprozesse durch eine 'entsprechend komplexe' Turingmaschine so wenig offensichtlich.

[20] Diese Sichtweise ist vereinbar mit Putnams Standpunkt, *Bedeutungen sind nicht 'im Kopf'* [Put91] Seite 137.
[21] Vergleiche Abschnitt 8.3 für eine ausführliche Diskussion des Komplexitätsaspektes.

7.4 Schlußfolgerungen für die KI

Aus der obigen Diskussion geht hervor, daß die phänomenologische Kritik von Dreyfus zwar nicht zeigt, daß Maschinen nicht Leistungen hervorbringen können, die menschlichen Intelligenzleistungen ebenbürtig sind.
Allerdings zeigt die phänomenologische Kritik einen für die *Entwicklung* intelligenter Systeme essentiellen Problembereich auf:
Ein großer Teil der algorithmischen Information in künstlichen intelligenten Systemen bezieht sich nicht auf die symbolische Speicherung der Verhältnisse einer materialen äußeren Welt. Aber worauf müssen dann die Symbole eines künstlich intelligenten Systems referieren ? Und welches ist die adäquate Beschreibungsebene für KI-Systeme ?
A. Newell schlägt eine sogenannte 'Wissensebene' (engl. *knowledge level*) in [New82] vor. Newell unterscheidet verschiedene Beschreibungsebenen von Computersystemen. Beginnend mit der physikalischen Ebene über eine Logik- und Registertransferebene gelangt er zur Symbolebene, welche den üblichen Programmiersprachen entspricht.
Die Wissensebene ordnet er oberhalb der Symbolebene an. Sie beschreibt das Systemverhalten durch Aussagen in einer zunächst nicht näher angegebenen Sprache. Die Aussagen zusammen mit der Annahme, daß das System sich 'rational' verhält, sollen gemeinsam das Verhalten bestimmen. Diese Wissensebene soll für den menschlichen Experten die Explikation seiner Vorgehensweise erleichtern. Der Terminus 'rational' ist hierbei allerdings nicht ganz unproblematisch - er kann in dieser allgemeinen Formulierung das Systemverhalten keineswegs eindeutig bestimmen.[22]
Lernverhalten, das induktives oder analoges Schließen beinhaltet, ist ein Beispiel für die Indeterminiertheit des Systemverhaltens durch das Rationalitätspostulat.[23] In der Tat ist es erforderlich, daß so etwas wie Präferenzrelationen auf konkurrierenden (Induktions-) Hypothesen spezifiziert werden. In solchen Präferenzrelationen drückt sich das implizite und - von Dreyfus als nicht repräsentierbar behauptete - Wissen aus, welches sich letztlich im faktischen Verhalten des Systems zeigt. Somit müssen also Symbole auf abstrakte Entitäten wie Präferenzrelationen verweisen. Es muß *das Verhalten des Systems*, wie es sich faktisch zeigen soll, symbolisch gespeichert werden. Dies bedeutet, daß die involvierten Symbole nicht auf ontologische Entitäten einer physischen Welt referieren können, sondern auf abstrakte Entitäten, welche das faktische Verhalten bestimmen. Diese abstrakten Entitäten haben unter anderem die folgenden Rollen auszufüllen:

- Präferenzerelationen zwischen konkurrierenden Hypothesen für die eindeutige Bestimmung von ansonsten unterbestimmten Schlußfolgerungen. Beispielsweise gibt es beim induktiven Schließen jeweils einander wiedersprechende Induktionshypothesen, die beide gleichermaßen durch die bisher bekannten Daten gestützt werden. Bei der Abduktion, bei der Entdeckung von Ursachen, kommen in der Regel eine Vielzahl von Möglichkeiten in Betracht. Das Vertrauen oder die Plausibilität, mit dem

[22] Es wird mittlerweile bezweifelt, daß die Wissensebene überhaupt ein nützliches Konstrukt ist. Siehe beispielsweise Schreiber et al. [SAW91] oder Vinkhuysen [Vin92].
[23] Vergleiche auch Dietterich [Die86].

die eine oder andere Möglichkeit für die Ursache gehalten wird, könnte durch solche Präferenzrelationen über mögliche Ursachen ebenfalls ausgedrückt werden.

- Im Fall des Einsatzes von sogenannten nicht-monotonen Schlußverfahren[24] muß aus Daten, die im logischen Sinn kein eindeutigen Schluß zulassen, doch irgendein Schlußergebnis bestimmt werden. Insofern liegt die Verwandtschaft des nicht-monotonen Schließens mit dem induktiven Schließen nahe.

- Bei der Handhabung des in Abschnitt 6.1 erwähnten Zeitkomplexitätsproblems sind Prioritätsregelungen erforderlich, um die Arbarbeitungsreihenfolge logischer Beweisversuche zu bestimmen. Dadurch wird letztlich auch das Verhalten bestimmt, da ein System im praktischen Einsatz irgendwann die Reihe von erfolglosen Beweisversuchen abbrechen muß, um seine Aktionen fortsetzen zu können. Dabei muß im Fall der erfolglosen Beweissuche irgendeine Entscheidung getroffen werden. Dies geschieht in der Regel durch die oben genannten nicht-monotonen logischen Schlußverfahren.

Da die phänomenologische Kritik darauf hindeutet, daß in diesen Bereich ein sehr großer Teil von algorithmischer Information einfliessen muß, wäre es konsequent, Ansätze zu entwickeln, die gerade die Akquisition von 'Wissen' dieser Art erleichtern. Diese Schlußfolgerung wurde erstmals in Hoffmann [Hof92] gezogen.[25]

[24] Eine Aufgabe für ein nicht-monotones Schlußverfahren ist beispielsweise: Alle Vögel außer den Pinguinen können fliegen. Karl ist ein Vogel. Kann Karl fliegen ? Diese Frage kann erst dann beantwortet werden, wenn geklärt ist, ob Karl ein Pinguin ist. Liegt diese Information nicht vor, so mag es vernünftig erscheinen anzunehmen, daß Karl fliegen kann, weil es weit weniger Pinguine als andere Vögel gibt. Da dieser Schluß später durch neue Informationen eventuell wieder zurückgenommen werden muß, wird ein solches Schlußverfahren auch *nicht-monoton* genannt.

[25] Experten werden mit dieser Art von Wissen allerdings erheblich größere Probleme haben, als beispielsweise mit der Explikation ihrer Kategorien für die jeweiligen Bereichsphänomene, z.B. mit der Angabe von Taxonomien ihres Fachgebiets. Letztere sind teilweise direkt aus einschlägigen Fach- und Lehrbüchern zu entnehmen.

Kapitel 8

Begriffe, Komplexität und Bewußtsein

Welche Rolle spielen Begriffe bei der menschlichen Erkenntnis ? Daß wir Menschen uns Begriffen bedienen, um Erkenntnis zu erlangen, ist unumstritten. Doch das Bild vom Ursprung und Wesen der Begriffe hat sich im Lauf der Jahrhunderte gewandelt.
Während der Empirismus die Herkunft der Begriffe aus der Erfahrung erklärte, sah der Rationalismus die Quelle der Begriffe im menschlichen Geist.[1]
Kant leitete die 'kopernikanische Wende' ein, und 'vereinigte' beide Sichtweisen: *Anschauungen ohne Begriffe sind blind.* Und *Begriffe ohne Anschauungen sind leer.* In diesem Jahrhundert wurde insbesondere durch Wittgenstein und Quine die anscheinend nur holistisch zu verstehende Bedeutung von Begriffen deutlich. Nicht nur in der Wissenschaftstheorie, sondern auch durch die Entwicklungen der künstlichen Intelligenz wurde die Begriffsproblematik von einem rein formalen Gesichtspunkt her wichtig; formal, d.h. in Absehung eines Bewußtseins, das sich der Begriffe bedient, um Erkenntnis zu erlangen. In KI-Systemen werden Symbole mechanisch manipuliert, die auf Begriffe verweisen.

Im folgenden werden zunächst neuere erkenntnistheoretische Sichtweisen von Wittgenstein (Abschnitt 8.1) und Quine (Abschnitt 8.2) vorgestellt. Die beiden Philosophen begannen ihre Untersuchungen auf der Ebene der faktischen menschlichen Begriffsverwendung.
In Abschnitt 8.3 wird umgekehrt ein System betrachtet, das ein sehr komplexes Verhalten zeigt. Vom Systemverhalten ausgehend, wird analysiert, inwiefern etwas zu unseren menschlichen Begriffen Korrespondierendes in solchen Systemen gefunden werden kann bzw. muß.
Abschnitt 8.4 und 8.5 befassen sich mit der Komplexität von künstlichen neuronalen oder konnektionistischen Systemen, bzw. mit der Fähigkeit von Systemen zur Selbstorganisation von komplexen Strukturen. Abschnitt 8.6 behandelt menschliche Kreativität und die mögliche Übertragung der Kreativität auf Maschinen. Es wird sich zeigen, daß auch für den (gängigen) Kreativitätsbegriff das menschliche Bewußtsein eine tragende Rolle spielt.

[1] Vergleiche auch Abschnitt 5.1, in dem ein Abriß der Geschichte der Allgemeinbegriffe gegeben ist.

8.1 Wittgensteins Regelbegriff

Wittgenstein untersuchte in seinen *Philosophischen Untersuchungen* [Wit53] die Bedeutung sprachlicher Ausdrücke, die er dort auf den *Sprachgebrauch* zurückführt. Wittgenstein betont, daß wir die Bedeutung von Worten durch Beispiele lernen. Wir lernen Begriffe oder Wortbedeutungen so, daß der jeweilige Geltungsbereich nicht scharf abgegrenzt ist. Wenn scharfe Grenzen von Begriffen angegeben werden können, so ist das die Ausnahme.
Regeln oder Begriffe werden nicht durch die logische Summe ihrer Beispiele explizit definiert. Die Regel oder der Begriff wird durch eine nicht weiter hintergehbare *Ähnlichkeit* gegeben, die unter den zu ihrer Definition angegebenen Beispielen besteht. Diese Ähnlichkeit wird durch jedes neu hinzukommende Beispiel fortgeschrieben, verengt, verändert oder auch erweitert.[2] Somit liegt ständig eine umfassende Menge von fragmentarischen Sprachverwendungs- bzw. Begriffsverwendungsregeln vor.
Damit erfolgt die Subsumption einer neuen Erfahrung, eines neuen Beispiels unter einer bestimmten Regel aufgrund erst fragmentarisch angeeigneter Regeln und Kategorisierungen. Nicht nur die fragmentarische Form der subsumierenden Regel wirkt bestimmend, sondern auch die fragmentarische Form möglicher Alternativregeln, unter denen die Erfahrung subsumiert werden könnte. Die Subsumption gründet damit in der nur holistisch zu begreifenden Gesamtheit der Regelfragmente. Die Gesamtheit der Regeln ist eingebettet in die menschliche Lebenspraxis. Die Kategorisierungen nehmen somit die Lebensgeschichte des Individuums in sich auf.

Das Erlernen von Regeln beinhaltet daher immer zwei Punkte:

- Einerseits die Erkennung ihrer Anwendungskriterien in konkreten Situationen, d.h. die Subsumption einer Erfahrung unter der jeweiligen Regel.

- Andererseits die spezifische *Fortschreibung*, die Veränderung, die Verengung oder Erweiterung der Regel aufgrund jeder neuen Erfahrung.

Dazu unterscheidet Quine die *Rezeptionsähnlichkeit* und die *Wahrnehmungsähnlichkeit*.[3] Das erste meint eine rein physikalische Ähnlichkeit der Einwirkung auf die Oberflächen der Sinnesorgane, ohne Rücksicht auf die kognitive Wirkung oder das Verhalten. Letzteres hingegen meint *keine* Ähnlichkeit der Reizeinwirkungen auf die Sinnesorgane, sondern eine Ähnlichkeit auf einem von dieser unmittelbaren physikalischen Ähnlichkeit weit abstrahierten konzeptuellen Ebene. Sie kann beliebig eingeschränkt sein, auf nur einen ganz bestimmten Aspekt und hängt von der *Disposition des Individuums ab*,[4] also von dem jeweiligen situativen Zusammenhang, in dem eine Erfahrung steht.
Daraus folgt für Wittgenstein, daß die Gesamtheit von Begriffsbedeutungen im menschlichen Denken und sprachlichem Handeln, konstitutiv für die menschliche Lebenswelt ist.

[2] Man denke beispielsweise an Wittgensteins *Familienähnlichkeiten*, bei denen immer neue Parallelen zwischen den zu subsumierenden Beispielen auftauchen.

[3] In Quine [Qui89] Seite 34.

[4] In Quine [Qui89] Seite 35.

Mithin bedeutet eine Sprache gemeinsam zu haben, auch eine Lebenswelt gemeinsam zu haben. Fachsprachen, Nationalsprachen, Slang oder Sprachen unter alten Freunden entstanden je unter spezifischer Praxis im Sprachgebrauch und sind jeweils *wirklichkeitskonstitutiv* für die einzelnen Teilnehmer der genannten Sprachpraxis.
Im Zusammenhang mit der künstlichen Intelligenz ist zu bemerken, daß Wittgensteins Untersuchungen zum Regelfolgen[5] deutlich machen, daß sich der Kern des menschlichen Regelfolgens einer exakten Beschreibung immer wieder entzieht. Beispiele dienen immer nur als (partielle) Explikation; das Verständnis der Regelhaftigkeit jedoch muß in irgendeiner Weise weiter reichen als alle Beispiele:

> 208. ... Es ist zu unterscheiden: das 'usw.', das eine Abkürzung der Schreibweise ist, von demjenigen, welches dies *nicht* ist. Das 'usw. ad inf.' ist *keine* Abkürzung der Schreibweise. Daß wir nicht alle Stellen von Π anschreiben können, ist nicht eine menschliche Unzulänglichkeit, wie Mathematiker manchmal glauben.
>
> Ein Unterricht, der bei den vorgeführten Beispielen stehen bleiben will, unterscheidet sich von einem, der über sie '*hinausweist*'.
>
> 209. 'Aber reicht denn nicht das Verständnis weiter als alle Beispiele ?' - Ein sehr merkwürdiger Ausdruck, und ganz natürlich ! -
>
> Aber ist das *alles* ? Gibt es nicht eine noch tiefere Erklärung; oder muß nicht doch das *Verständnis* der Erklärung tiefer sein ? *Habe* ich mehr, als ich in der Erklärung gebe ? - Woher aber dann das Gefühl, ich hätte mehr ?
>
> Ist es, wie wenn ich das nicht Begrenzte als Länge deute, die über jede Länge hinausreicht ? [6]

Für unsere Zwecke lassen sich die zwei folgenden Aspekte grundsätzlich voneinander unterscheiden:

1. Der bisherige Geltungsbereich einer Regel; welcher gewöhnlich durch Beispiele ansatzweise expliziert wird.
2. Die Art und Weise wie sich eine Regel *fortschreibt*; welche sich immer nur im Rückblick auf vergangene Veränderungen der Regelanwendungen bestimmen läßt.

Es läßt sich feststellen, daß Wittgensteins Untersuchungen keine *operationale* Erklärung bzw. Beschreibung der menschlichen Regelverwendung und insbesondere der Regel*fortschreibung* geben. Vielmehr sind es Beschreibungen auf einer Metaebene. Es wird darauf verwiesen, daß es noch bestimmte nicht weiter beschriebene Prozesse gibt, die letztlich die Regelfortschreibung im Einzelnen bestimmen. Insofern können die Überlegungen Wittgensteins zum Regelbegriff in der menschlichen Begriffsverwendung noch nicht unmittelbar für die Entwicklung von KI-Systemen genutzt werden.

[5] Man vergleiche auch die neuere Diskussion zu Wittgensteins Begriff des Regelfolgens, ausgelöst durch Kripke [Kri82]. Siehe beispielsweise Blackburn [Bla84], Pettit [Pet90] und Pears [Pea91].

[6] Wittgenstein [Wit53] §209

8.2 Quines Bedeutungsholismus

W.v.O. Quine wies darauf hin, daß es keinen Sinn hat, für einen einzelnen Satz feststellen zu wollen, ob er wahr oder falsch ist. Der Grund dafür liegt in dem Problem, daß die Geltungsbestimmungen eines einzelnen Satzes praktisch immer eine ganze Reihe von anderen Sätzen als bereits geltend voraussetzen. Wenn ein bestimmter Satz auf den ersten Blick nicht zu gelten scheint, so hat man bei näherem Hinsehen eine Menge von insgesamt einander widersprechenden Sätzen.
Nun stellt sich die Frage, welcher dieser Sätze zu revidieren ist. Und dies läßt sich nicht durch *logische* Mittel allein entscheiden. Dies hängt mit der nur holistisch zu erfassenden Bedeutung jedes einzelnen Satzes zusammen.

Gegeben seien beispielsweise die folgenden vier Sätze:

1. Alle Vögel können fliegen.
2. Karl ist ein Pinguin.
3. Alle Pinguine sind Vögel.
4. Karl kann nicht fliegen.

In den obigen vier Sätzen ist ein Widerspruch enthalten. Nun stellt sich die Frage, welcher oder welche der vier Sätze revidiert werden muß bzw. müssen, um den Widerspruch aufzulösen. In jedem Fall läßt sich beobachten, daß man durch geeignete Abänderung jedes einzelnen der vier Sätze die Konsistenz der Gesamtheit der Sätze herstellen kann. Daher gibt es also keinen *logischen* Grund für die Abänderung gerade eines bestimmten Satzes, um die Widerspruchsfreiheit herzustellen.
Nur eine Theorie als Ganzes kann verworfen werden. Jeder einzelne Satz hingegen kann bei geeigneter Anpassung der übrigen Theorie akzeptiert werden.
Quine bezieht sich hierbei nicht nur auf die empirischen oder synthetischen Sätze, er sieht auch analytische Sätze, Sätze der Logik etwa, als reversibel an. Verlangt man zur Festlegung der *Bedeutung* eines einzelnen Satzes die Angabe von dessen *empirischen Geltungsbedingungen*, so läßt sich aus dem oben Gesagten ableiten, daß auch die Bedeutung eines einzelnen Satzes nicht für sich festgelegt werden kann, sondern immer nur vor dem Horizont einer umfassenden Theorie des jeweiligen Gegenstandsbereiches zu sehen ist.[7]
So schreibt Quine zu den Sätzen einer empirischen Theorie:

> ... sie liegen an der Peripherie, wo ihre Bedeutung empirisch durch Aufforderung zu Zustimmung oder Ablehnung in jeder einzelnen Situation festgestellt werden kann. Von dieser Peripherie empfängt die Wissenschaft und die Sprache ihren gesamten empirischen Gehalt oder ihre Bedeutung. Um dagegen die Bedeutung eines Satzes von ewiger Dauer tief im Inneren der Theorie herauszuschälen, kann man sich nur auf seine vielfältigen Verbindungen innerhalb

[7] Vergleiche Putnam [Put86].

> der Theorie und letzten Endes, indirekt, mit der Peripherie stützen. Da jeder solche Strang nur mittels seiner Verschränkungen mit anderen beschreibbar ist, verliert die Frage nach der Bedeutung eines einzelnen solchen Satzes jeglichen Sinn. Der Satz könnte auch durch andere seiner Art formuliert werden, und vielleicht kann ein umfangreicheres Netz solcher Sätze eine gemeinsame Erklärung anhand ihrer effektiven Gesamtrelevanz für Beobachtungen und Situationssätze erfahren.[8]

Die Quinesche Auffassung kommt insofern Wittgenstein sehr nahe. Wittgenstein hat sich jedoch auch sehr intensiv gerade mit dem dynamischen Aspekt des Regelerlernens und der Regelfortschreibung befaßt.

8.3 Begriffe in komplexen Strukturen

Welche Rolle können oder müssen Begriffe oder sprachlichen Ausdrücke auf der Folie sehr hoher Strukturkomplexität eines kognitiven Systems spielen ? Kann durch Begriffe zumindest zum großen Teil das resultierende Verhalten eines Individuums erklärt werden ? Um diese Fragen zu klären, ist zunächst der Bezug zu klären, den Begriffe, sprachliche Ausdrücke oder einfache Aussagesysteme in Hinblick auf resultierendes Verhalten haben können bzw. müssen.

8.3.1 Komplexes Verhalten und dessen kompakte Organisation

Die Aufgabe eines komplexen Systems in einer natürlichen Umwelt stellt sich derart, daß es sich zunächst einmal nur in irgendeiner Weise 'sinnvoll' verhalten muß. Dazu zählt gegebenfalls auch ein geeignetes Sprachverhalten.
Das System muß sich seiner Umwelt in gewisser Weise anpassen. Zum Beispiel, falls Nahrung nicht an dem gewohnten Ort zu finden ist, Nahrung anderweitig zu besorgen.
Hier ist also eine geeignete Veränderung des üblichen Verhaltens erforderlich. Bei sprachlichen Äußerungen würde diese Forderung erheblich über die sonst üblichen Randbedingungen des einfachen Lebens hinausgehen. Dann ist auch das geeignete Verwenden sprachlicher Ausdrücke sowie angemessene Reaktionen auf sprachliche Äußerungen anderer Individuen gefordert.
Damit ein System sich an seine Umwelt anpassen kann, muß überhaupt erst einmal ein geeignetes Verhaltensrepertoire zur Verfügung stehen. 'Anpassung' meint hier, im Falle einer Veränderung der Umwelt, das Systemverhalten angemessen zu verändern.
Das System muß also für die faktisch auftretenden äußeren Situationen über je geeignete Verhaltensweisen verfügen.
Der Begriff der *Änderung* (oder Anpassung) einer Verhaltensweise an eine Umwelt hängt dabei nicht nur von der Umwelt ab, sondern auch von dem, was als Verhaltensweise gilt,

[8] Aus Quine [Qui89] S. 95.

und damit auch von den impliziten Erwartungen an die Gegebenheiten in der Umwelt.[9]
Im folgenden sollen unter *Verhaltensweisen* nicht nur stereotype Bewegungen verstanden werden, sondern auch Verhalten, dem komplexere Strukturen zugrunde liegen. Eine verhältnismäßig einfache Struktur könnte beispielsweise das *Zähneputzen* betreffen. Dort könnte eine einfache Verhaltensweise, der periodischen Bewegung der Zahnbürste auf der Oberfläche der Oberkieferzähne entsprechen. Eine umfassendere und damit auch komplexere Verhaltensweise könnte das Putzen des gesamten Gebisses einschließlich der Vor- und Nachbereitung beinhalten. Darin wäre das Auffinden der Zahnbürste inbegriffen, die an je verschiedenen Orten liegen kann. Wenn die Zahnbürste sich etwa noch innerhalb des Badezimmers befindet, so würde man vielleicht kaum von einer Anpassung an die Umwelt sprechen. Ist die Zahnbürste hingegen dort nicht zu finden, so kommt schon eher der Terminus *Verhaltensänderung* für das Systemverhalten in Betracht - beispielsweise weil auf das Zähneputzen ganz verzichtet wird, oder erst eine Zahnbürste eingekauft wird.
Der Fall in dem eine Verhaltensänderung erforderlich wird, könnte man auch mit Heideggers *Störung des Verweisungszusammenhangs* vergleichen. Durch Heideggers Begriff der *Störung* liesse sich damit auch der Begriff der *Verhaltensweise* definieren. Bei Heidegger ist diese Begrifflichkeit jedoch eng mit einem involvierten Bewußtsein und bewußt werdenden Störungen verbunden.
Da wir jedoch bei unseren Betrachtungen von einem Bewußtsein und damit auch von jedem möglichen Bewußten absehen, läßt sich der Begriff der Verhaltensweise nur durch bestimmte Merkmale seiner Beschreibung - seiner algorithmischen Beschreibung - charakterisieren.
Um eine unnötig umfangreiche Beschreibung des gesamten Verhaltensspektrums zu vermeiden, lassen sich Situationsklassen bilden, in denen jeweils das gleiche Verhalten gezeigt werden soll. Diese Situationsklassen lassen sich auf verschiedenen Hierarchieebenen bilden. Beispielsweise könnten die folgenden Situationsklassen auf je verschiedenen Hierarchieebenen gebildet werden:

1. Situationen in denen ein Arm gehoben werden soll.
2. Situationen in denen ein Gegenstand ergriffen werden soll, der sich irgendwo im Umkreis von wenigen Metern des Systems befindet.
3. Situationen, in denen eine Lebensmittelkonserve geöffnet werden soll.
4. Situationen in denen Nahrung beschafft werden soll.

Durch eine derartige hierarchische Klassifikation von Verhaltenssituationen ensteht eine Struktur, der man auf den verschiedenen Ebenen auch Ontologien, konkrete und abstrakte, Einzel- und Allgemeinbegriffe zuordnen kann - und müsste, würde man nur ein erkennendes Bewußtsein, das ähnlich dem menschlichen Bewußtsein ist, dabei voraussetzen.

[9] In diesem Sinn erwartet eine Verhaltensweise implizit, daß ihre faktische Ausführung 'erfolgreich' ist, und nicht - beispielsweise bei dem Griff nach einem Gegenstand buchstäblich ins Leere greift.

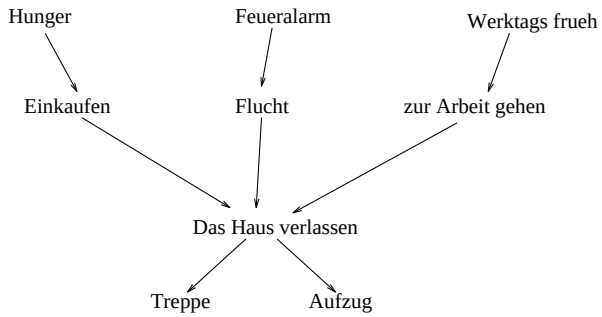


Abbildung 8.1: Beispiel einer hierarchischen Situationsklassifikation zur Strukturierung des Verhaltenspotentials: Verhaltensweisen, das Haus zu verlassen unterscheiden sich im wesentlichen dadurch, ob die Treppe oder der Aufzug benutzt wird. Dieser Unterschied ist allerdings irrelevant, wenn eingekauft oder zur Arbeit gegangen werden soll. Nur im Fall eines Feueralarms ist die Verhaltensstrategie 'Haus verlassen' nicht uneingeschränkt benutzbar.

Die obigen Beispiele von Situationsklassen beziehen sich zwar auf typisch menschliches und europäisch tradiertes Denken - allerdings ist festzuhalten, daß einer kompakten Beschreibung langer Zeichenketten notwendigerweise eine gegebenfalls hierarchische Klassenbildung entspricht.[10]
Bei der in Abschnitt 5.2 genannten Annahmen einer ausführlichen Beschreibung allen potentiellen menschlichen Verhaltens durch eine Zeichenkette der Länge $\approx 10^{10^{14}}$ und deren angenommene mögliche kompakte algorithmische Beschreibung durch eine Zeichenkette der Länge 10^{12} ist eine Bildung von umfangreichen Situationsklassen unvermeidlich. Wieviele Hierarchieebenen bei einer solchen notwendigen Klassenbildung entstehen *müssen*, hängt vom spezifischen Einzelfall der kompakt zu beschreibenden Struktur ab. Ob eine solche Klassenbildung typisch menschlichen Kategorisierungen entsprechen muß, ist damit ebenfalls noch nicht beantwortet.
Weiterhin läßt sich festhalten, daß die Klassenbildung durch eine bestimmte Vorgabe von Verhalten keineswegs eindeutig bestimmt wird. Siehe Abbildung 8.2 in der ein ein einfaches Beispiel dafür gegeben ist. Wie dem auch sei, eine Klassenbildung von Situationen, in denen je gleiches bzw. analoges Verhalten gefordert ist, wird sich an tatsächlich vorkommenden Situationen orientieren müssen. Die Klassenbildung muß also implizit oder explizit wiederspiegeln können, in welche *adäquate* Situationsklasse die augenblickliche Situation (Umwelt-System-Verhältnis) einzuordnen ist. Die Klassenbildung muß also die Grundlage dafür bieten, die Umweltsituationen adäquat zu unterscheiden. Es muß anzeigbar sein, ob bestimmte Situationscharakteristika *vorliegen oder nicht vorliegen*, die bestimmte Verhaltensweisen erfordern. Das heißt, die Welt (Umwelt) gliedert sich dadurch in Erfüllungsbedingungen der verschiedenen Situationsklassen auf.[11] Von der Vor-

[10] Es müssen bei einer Kompaktierung sich wiederholende Strukturen auch als solche beschrieben werden.

[11] Damit ist man im Grunde bei der Struktur angelangt, wie sie auch bei I. Kants Analyse des reinen Verstandes deutlich wird: *Begriffe ... die Einheit der Handlung verschiedene Vorstellungen unter einer*

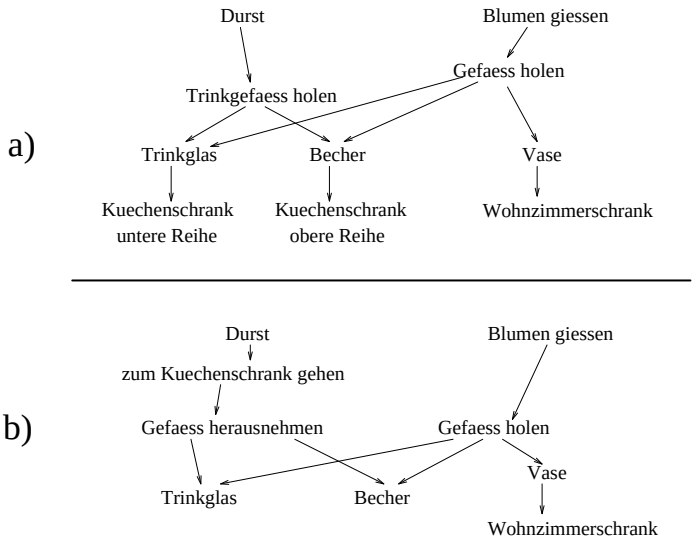


Abbildung 8.2: In obigem Schema sind verschiedene Kategorisierungen vorgenommen, die insgesamt zu äquivalentem Verhalten führen. Während in a) der eingeschränkte Begriff 'Trinkgefäß' erforderlich ist, um aus keiner Vase trinken zu müssen, wird dies in b) implizit dadurch gewährleistet, daß die Verhaltensstrategie vorschreibt, erst zu dem Küchenschrank zu gehen. In diesem ist faktisch keine Vase enthalten. Dadurch kommt das System ohne die entsprechende Unterscheidung aus, und zum gleichen Ziel.

aussetzung, daß ein System eine komplexe Verhaltensbeschreibung erfordert, gelangten wir zu einem ganz ähnlichen Ergebnis, was die potentielle Entstehung oder Erzeugung einer 'Weltvorstellung' angeht; jedoch fehlt bei unserer Analyse gänzlich eine erkenndendes Bewußtsein.

Somit wurde dargelegt, daß sich zwangsläufig eine hierarchische Klassenbildung - implizit oder explizit - bei der kompakten Organisation eines intelligenten Verhaltensrepertoires ergibt.
Die obigen Ausführungen gelten natürlich nicht nur für 'motorisches Verhalten', sondern auch für Sprachverhalten. Das Sprachverhalten würde sich ebenfalls entsprechend den Erfordernissen ergeben, die in der jeweiligen Lebensumgebung typischerweise mitgeteilt werden müssen. Damit erscheint es in hohem Maße kulturabhängig. Dies gilt entsprechend der obigen Ausführungen auch für die sich parallel entwickelnde 'interne Ontologie' eines solchen Systems.

gemeinschaftlichen zu ordnen. (Aus Kants Kritik der reinen Vernunft [Kan87] B 93.) Während Kant den menschlichen Erkenntnisapparat analysierte, sind wir sozusagen von der entgegengesetzten Seite gekommen.

8.3.2 Antropomorphe Begriffe in komplexen Strukturen

Wenn eine einfache Struktur - diese soll eine mögliche Erkenntnis repräsentieren, beispielsweise eine Beziehung zwischen Begriffen (z.B. Alle p sind q) - sich in einem letztlich komplexen Verhalten niederschlagen soll, so muß zunächst eine weitere komplexe Struktur vorhanden sein, in die die einfache Struktur eingebettet ist.
Weiterhin kann man allgemein sagen, daß die einfache Struktur entweder das tatsächliche Verhalten direkt beschreibt - dann kann es sich aber nur um einen sehr kleinen Teil des gesamten Verhaltensspektrums handeln.
Oder aber die Struktur hat ihre Bedeutung für einen umfangreichen Teil des Verhaltens - dann müssen aber die Bedingungen der konkreten Umsetzung von der einfachen Struktur - hier als theoretische Erkenntnis gedacht - in praktisches Verhalten bzw. Handeln entsprechend kompliziert sein - z.B. in Form von nur komplex zu explizierenden und in der Theorie nicht definierten Grundbegriffen.
Dies bedeutet ebenfalls, daß der Anwendungsbereich nur durch komplexe Regeln abzugrenzen ist. Die beiden genannten Fälle bilden die zwei äußeren Pole eines Spektrums von Möglichkeiten, innerhalb dessen die 'einfache Erkenntnis' irgendwo eingeordnet werden muß. Dies geht darauf zurück, daß keine, wie auch immer geartete, *einfache* Theorie das resultierende komplexe Verhalten erklären kann. Mithin kann sie nur - wenn sie überhaupt an der Bestimmung des resultierenden Verhaltens potentiell beteiligt sein soll - eingebunden sein in eine große Zahl von algorithmischen Regeln, die das letztlich resultierende Verhalten gemeinsam mit der 'einfachen Erkenntnis' bestimmen.
Beispielsweise könnte eine 'einfache Erkenntnis' beschreiben, wie das Wort *Tisch* geschrieben wird. Dies wird jedoch nur einen verschwindend geringen Teil des Gesamtverhaltens eines intelligenten Systems ausmachen.
Auf der anderen Seite könnte die 'einfache Erkenntnis' die sinnlichen Erkennungskriterien für einen *Tisch* beschreiben. Diese Erkenntnis könnte dann nicht nur dafür genutzt werden, im richtigen Moment *Tisch* zu schreiben, sondern in einer Vielzahl von Situationen etwas als einen Tisch zu erkennen und beispielsweise ihn entsprechend zu benutzen, bzw. allein die Möglichkeiten des Benutzens in verschiedene Pläne und Handlungen einfliessen zu lassen.

Eine zweite Dimension innerhalb dessen sich 'einfache Erkenntnis' bewegen kann, ist ihr mehr oder weniger direkter Einfluß einerseits auf die 'Datenaufnahme' und deren Strukturierung und Einordnung, und andererseits ein mehr oder weniger direkter Einfluß auf das resultierende Verhalten, die 'Datenausgabe'. Siehe dazu Abbildung 8.3.
Im allgemeinen befassen wir uns wohl mit Erkenntnissen, die mehr oder weniger in der Mitte beider Dimensionen des genannten Spektrums liegen. Wenn wir uns mit Begriffsklärungen, z.B. dem Begriff des *Tisches* befassen, so liegen sie in folgender Weise in der Mitte:
Im Hinblick auf den Umfang des mehr oder weniger indirekten Einflusses auf das resultierende Verhalten, ist ein solcher Begriff von einigermaßener Breite in seiner Anwendung, jedoch deutlich weniger breit, als etwa allgemeine Handlungsmaximen.
Innerhalb der zweiten Dimension, der Ferne oder Nähe zur unmittelbaren Ordnung von

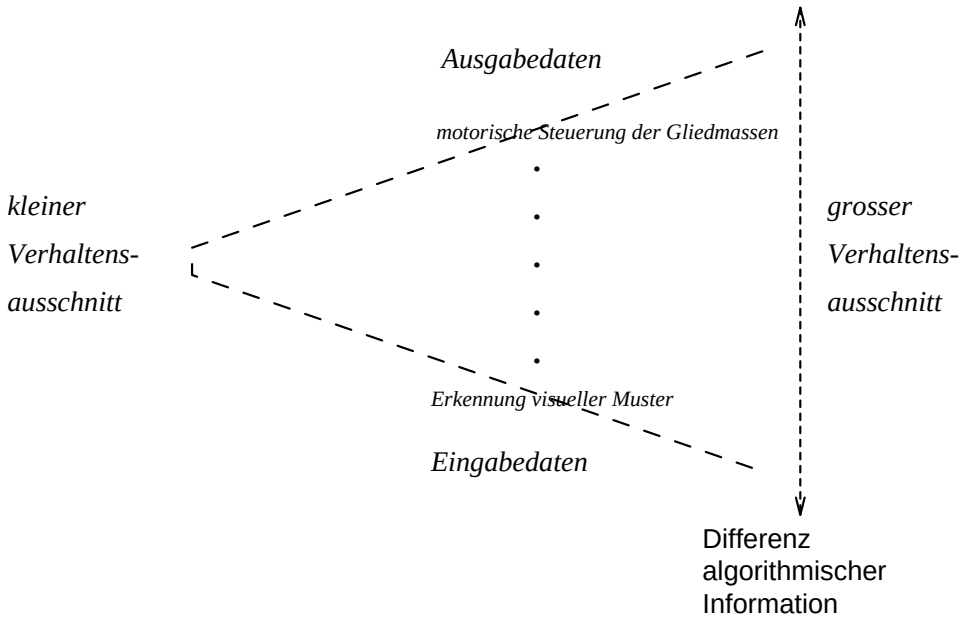


Abbildung 8.3: Zweidimensionales Spektrum innerhalb dessen sich 'einfache Erkenntnis' ansiedeln kann.

'Eingangsdaten' bzw. 'Ausgangsdaten' stellt er ebenfalls kein Extrem dar. Von den unmittelbaren 'Eingangsdaten'[12] ist er durch eine Vorverarbeitung dieser Daten zu abstrakteren Einheiten (z.B. gerade Linien bei der visuellen Wahrnehmung) entfernt. Von den unmittelbaren 'Ausgangsdaten' ist er durch Nachverarbeitungen entfernt, die erforderlich sind, um zu einem spezifischen Verhalten zu gelangen - aufgrund der Subsumption einer gegebenen Wahrnehmung unter dem Begriff *Tisch*.
Wird nun gefordert, einen solchen irgendwo im mittleren Bereich des Spektrums liegenden Begriff durch eine 'handliche' Definition zu explizieren, beispielsweise durch andere Begriffe aus der gleichen Region des Spektrums, so stellt sich das folgende Problem:
Es sollen hier immer nur 'einfache' Explikationen in Betracht gezogen werden. Das heißt, die geschilderten Beziehungen zur Erklärung des Begriffs *Tisch* müssen kurz sein.
Wenn eine solch 'einfache' Beziehung zwischen den involvierten Begriffen tatsächlich besteht, so müssen notwendigerweise die Erfüllungsbedingungen für die definierenden, für die erklärenden Begriffe, entsprechend komplex sein.
Falls diese ebenfalls einfach sein sollten, muß der gesamte Begriffskorpus entsprechend 'weit' von den 'Ausgabedaten', dem Verhalten, entfernt sein. Dies läuft auf die Denkfigur hinaus, die auch Quines Bedeutungsholismus zugrunde liegt.

8.3.3 Beschränktes Bewußtsein als Fenster zur Komplexität

Wie kann Introspektion durch ein Bewußtsein geschehen, das nur ein stark beschränktes Fassungsvermögen hat ?

[12] Beispielsweise visuelle oder taktile Wahrnehmungen.

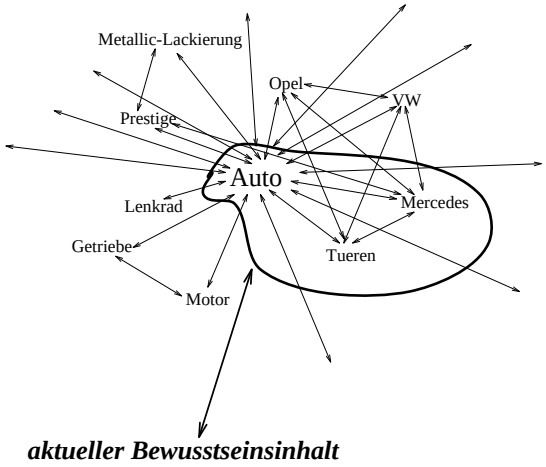


Abbildung 8.4: Begriffsnetz und beschränktes Bewußtsein.

Wenn das Verhalten eines kognitiven Systems von sehr hoher Komplexität ist, das Medium mittels dessen diese Struktur analysiert werden soll, aber nur einen verschwindend geringen Teil dieser Struktur gleichzeitig fassen kann, so stellt sich nicht nur die Frage nach den generellen Möglichkeiten und Grenzen der Analyse, sondern auch nach vielleicht typischen Charakteristika möglicher Vorgehensweisen.
Wie im vorherigen Abschnitt angerissen, stellt sich eine komplexe Struktur - auch wenn es sich nur um einen kleinen Ausschnitt der Struktur handelt - im allgemeinen wie folgt in einem beschränkten 'Fenster' dar: Siehe hierzu Abbildung 8.4.
In dem Fenster können nur ein oder eventuell einige wenige Begriffe, d.h. so etwas wie symbolische Darstellungen dessen, enthalten sein. Jedoch haben sehr viele Begriffe notwendigerweise eine sehr große Zahl von Bezügen zu anderen Begriffen bzw. zu anderen Teilen der komplexen Struktur. Nicht nur die Begriffe, zu denen ein jeweils bewußt gemachter Begriff Bezüge hat, können nicht alle gleichzeitig bewußt sein, sondern auch die Bezüge selbst können höchstens zu einem Bruchteil gleichzeitig im Bewußtsein vorliegen. Beschränkt man sich bei dieser Betrachtung nun auf Einzelbegriffe, die sich auf singuläre Gegenstände beziehen, so wird deutlich, daß nicht alle in einer aktuellen Handlungssituation äußerlich präsenten und relevanten Gegenstände bewußt sein können.
Somit muß aufgrund der Beschränktheit des Bewußtseins ständig eine erhebliche Menge handlungsrelevanter Gegenstände lediglich *zuhanden* sein, um Heideggers Terminologie zu verwenden. *Vorhanden* kann allenfalls ein Bruchteil der relevanten Entitäten sein. Erst wenn die 'alltäglichen Verweisungen' gestört werden, muß das komplexe System nach einer Verhaltensänderung suchen. Dafür orientiert sich das System an verhaltensrelevanten Klassifikationen, die sich sinnvollerweise an der akuten Störung orientieren. Die vorliegende gegenständliche Situation, das *Zeug* wechselt seine *Zuhandenheit* zu einer *Vorhandenheit*.
Insofern ist zwar a priori nicht auszuschließen, daß sukzessive die gesamte kognitive Struktur reflektiert und explizit gemacht werden kann - ob dies möglich ist, würde von weiteren

Eigenschaften des 'Fensters' abhängen.[13] Jedoch wird deutlich, daß es keinen Sinn haben kann, Begriffe im allgemeinen durch einfache Strukturen erklären oder definieren zu wollen. Prinzipiell könnten einige Begriffe zwar so definiert werden - z.B. Synonyme[14] - aber für den größten Teil der Begriffe muß eine Klärung auf eine umfangreiche, eine komplexe Struktur zurückgreifen.
Dies ergibt sich aufgrund der Komplexität des gesamten (kognitiven) Systems. Wie sich einzelne Situationsklassen bzw. Begriffe bilden und fortschreiben, ist von dieser Analyse noch unberührt. Unter anderem deutet Heideggers Phänomenologie ebenso wie Wittgensteins Philosophische Untersuchungen darauf hin, daß dies selbst wiederum ein komplexer Prozeß (auch im Sinne der Kolmogoroffkomplexität) ist. Dies könnte bedeuten, daß der Mensch nicht durch Introspektion, insbesondere nicht durch Reflexion anschaulicher Begrifflichkeiten, diese Prozesse erfassen kann. Daraus würde sich ein Problem bei der Wissensakquisition ergeben, das - wie bereits in Abschnitt 7.4 angedeutet - den Experten gleichermaßen wie den Wissensingenieur vor ungeahnte Probleme stellt.

8.3.4 Schlußfolgerungen

Wittgenstein untersuchte in seinen philosophischen Untersuchungen [Wit53] die menschliche Begriffs- und Sprachverwendung. Dort kam er ganz ähnlich wie Quine zu dem Ergebnis, daß Begriffe im allgemeinen nicht durch einfache Definitionen vollständig erklärt werden können.
Quine wies darauf hin, daß einzelne Aussagen ihre Bedeutung und damit ihren empirischen Wahrheitsgehalt letztlich aus der Bedeutung eines gesamten Netzes von Begriffen beziehen; daß Satzbedeutungen nur holistisch in Beziehung zu dem Rest des Wissens bzw. der für wahr gehaltenen Aussagen zu erfassen sind.
Die vorliegende Analyse in Abschnitt 8.3.1 und 8.3.2 eines komplexen Systems und einer möglichen Identifizierung von Strukturelementen mit menschlichen Begriffen zeigte, daß bereits aus den Annahmen

1. einer sehr hohen Strukturkomplexität und

2. daß Begriffe in einfachen Verhaltenserklärungen effektiv verwendet werden können,

folgt, daß einzelne Begriffe letztlich nur holistisch zu erfassen sind, wie es Quine aufgrund von Begriffsanalysen erkannte.

Im Gegensatz zu Heidegger sind wir zunächst von einem komplexen System ausgegangen, und fragten nach den Möglichkeiten, die ein Bewußtsein hat, das einerseits nur ein sehr eingeschränktes Fassungsvermögen besitzt und andererseits die Bewußtseinsinhalte doch

[13] Dies würde z.B. davon abhängen, ob tatsächlich alle wesentlichen Wissensstrukturen bewußt gemacht und kohärent expliziert werden können. Man denke beispielsweise an die Aufbereitung elementarer visueller Eindrücke.

[14] Quine bezweifelt, daß es so etwas wie Synonyme wirklich gibt. Siehe dazu Quine [Qui80]. Eine kritische Diskussion diser These findet sich in von Kutschera [von80].

einen starken Bezug zu dem resultierenden Verhalten haben sollen. Ausgehend von diesen Prämissen gelangten wir zwangsläufig zu einer Unterscheidung von *Zuhandenheit* und *Vorhandenheit*, dem *Zeug* und einer (relativen) Ontologie, die sich aus der Störung des Verweisungszusammenhangs des alltäglichen Handelns, der Sorgestruktur ergibt.

8.4 Komplexität in konnektionistischen Systemen

Wie bereits in Abschnitt 2.8 kurz dargestellt, haben konnektionistische Systeme in den letzten Jahren erheblich an Popularität gewonnen. Dreyfus [DD87] weist beispielsweise darauf hin, daß konnektionistische Verarbeitungsmodelle eine frappante Ähnlichkeit zu der scheinbaren Entstehungsweise menschlicher Kognitionen zeigen. Die Biologen Maturana und Varela haben eine neue Theorie über die Entstehung von biologischen kognitiven Systemen entwickelt, die sie auf empirische Forschungsergebnisse aus der Biologie stützen [MV87, Var90]. Darauf wird im nächsten Abschnitt noch näher eingegangen.
Dreyfus, Smolensky und andere sehen im Konnektionismus einen vielversprechenden Weg zu neuen Modellen menschlicher Intelligenz. Auch für die Kognitionswissenschaft wird den konnektionistischen Modellen große Chancen eingeräumt.[15] Konnektionistische Systeme basieren auf der Idee, eine Vielzahl von Verarbeitungseinheiten in einem System zu haben. Die Verarbeitungseinheiten sind untereinander stark vernetzt. Dadurch beeinflussen sich die einzelnen Verarbeitungseinheiten ständig gegenseitig in ihrer Funktion. Insbesondere erhofft man sich bei konnektionistischen Rechnerarchitekturen eine besondere Lern- bzw. Anpassungsfähigkeit in einer unbekannten Umgebung, die herkömmlichen Rechnerarchitekturen überlegen sein soll.[16]
Ein anderer Grund für die großen Hoffnungen, die in konnektionistische Ansätze gesetzt werden, mag die außerordentliche Schwierigkeit der exakten Analyse von dem sein, was in einem großen Netzwerk tatsächlich passiert und was potentiell passieren kann, nachdem eine Reihe von Eingaben stattgefunden haben. Wegen dieser Schwierigkeit kann man auch nicht so leicht erkennen, ob oder welche konnektionistischen Ansätze zu verwerfen sind. Minsky und Papert [MP69] analysierten sogenannte Zweilagennetze auf ihr potentielles Lernverhalten. Dort stellten sie schwerwiegende Beschränkungen der Berechnungs- und Lernfähigkeit solcher Netze fest. Als Folge wurde die Förderung vieler Forschungsprojekte zu neuronalen Netzen gestrichen. Aber selbst die Analyse von solch eingeschränkten Netzen hatte bereits einen außergewöhnlichen Schwierigkeitsgrad, wie Papert [Pap88] betont. Und die Analyse von komplizierteren Netzen erscheint noch erheblich schwieriger zu sein.
Für die Einschätzung der generellen Möglichkeiten von neuronalen oder konnektionistischen Berechnungsmodellen wurde von mir in Hoffmann [Hof90b] der Begriff der Kolmogoroffkomplexität vorgeschlagen. Der Vorteil dieser Methode besteht darin, daß eine exakte Analyse der einzelnen Netzwerkelemente und das teilweise hochkomplizierte Inein-

[15] Siehe beispielsweise Varela [Var90] oder Smolensky [Smo87, Smo88].

[16] Ein neuerer Überblick über die technischen Details der verschiedenen Ansätze zum Lernen in konnektionistischen Systemen ist in Hinton [Hin89] zu finden.

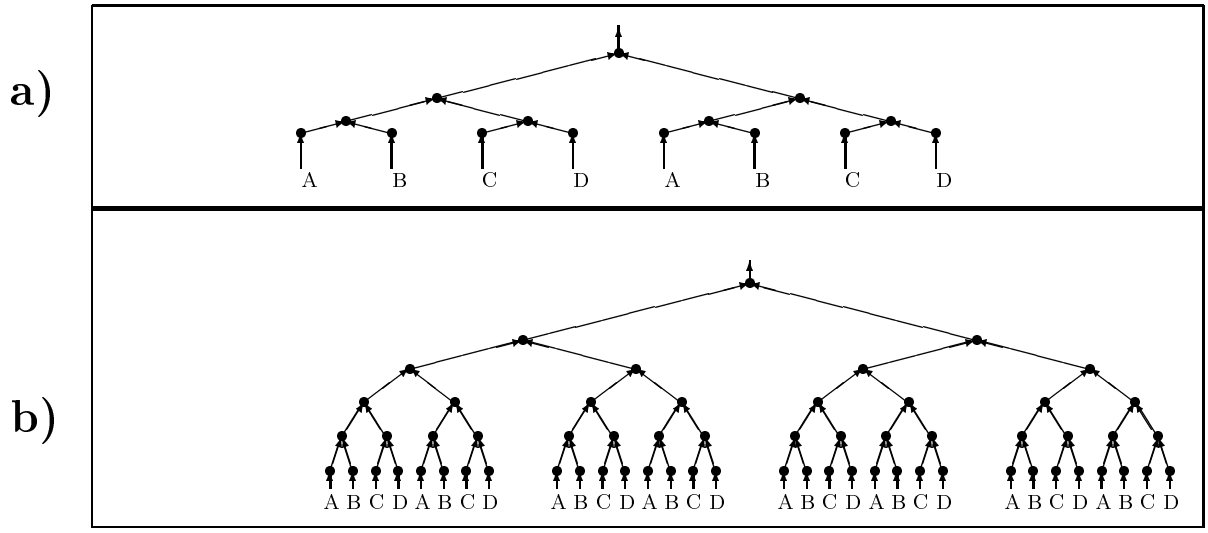


Abbildung 8.5: Zwei unterschiedlich große Netzwerke ähnlicher Struktur.

andergreifen der vielen lokalen Knotenaktivitäten nicht erforderlich ist, um Aussagen über die mögliche Komplexität der letztendlich resultierenden (emergenten) Ausgangssignale zu machen.
Die Grundidee besteht darin, daß eine Beziehung zwischen der Entwurfs- oder Beschreibungskomplexität des Gesamtnetzwerkes und der Komplexität der damit berechenbaren Funktionen hergestellt wird.
Eine solche Beziehung wurde auch für lernende Netzwerke in Hoffmann [Hof90b] hergestellt. Dabei werden die möglichen Veränderungen der internen Zustände jedes einzelnen Netzwerkknotens erfaßt.
Aufgrund dieser Betrachtungen läßt sich sehr leicht sagen, daß beispielsweise das Netzwerk b) in Abbildung 8.5 keine wesentlich komplexere emergente Funktion berechnen kann, als das wesentlich kleinere Netzwerk a). Dies läßt sich daran erkennen, daß eine Beschreibung des Netzwerkes b) nicht wesentlich umfangreicher sein muß, als die Beschreibung des Netzwerkes a). In Netzwerk b) ist das Netzwerk a) mehrfach enthalten.
Als Konsequenz dieses Zusammenhangs konnte in Hoffmann [Hof91b] gezeigt werden, daß generell die Bedeutung der Funktionalität des einzelnen Netzwerkknotens in einem großen Netzwerk gegenüber der Netzwerktopologie von immer geringerer Bedeutung für die Komplexität der emergenten Ausgangssignale ist.[17]

[17] M. Arbib weist in Arbib [Arb89] auf Seite 51 darauf hin, daß "there is no such thing as a 'typical' neuron." Das heißt, daß das menschliche Gehirn, abgesehen von seiner immensen Neuronenanzahl und der riesigen Zahl von unregelmäßigen Verbindungen auch noch sehr viele verschiedene Neuronen besitzt, wodurch die Beschreibungskomplexität nochmals gesteigert wird. Man könnte zwar Abschätzungen über den möglichen Informationsgehalt machen - jedoch wird es dabei im Dunkeln bleiben, inwieweit die

Wenn man sich Wittgenstein und Quine zur äußeren Erscheinung der menschlichen Sprach- bzw. Begriffsverwendung noch einmal verdeutlicht, so sind offensichtliche Parallelen zu den in den letzten Jahren so populär gewordenen subsymbolischen Ansätzen zur Informationsverarbeitung zu finden.[18] Auch in vielen konnektionistischen Verarbeitungsmodellen lassen sich die folgenden Eigenschaften beobachten:

a) Die Subsumption einer Erfahrung (eines Eingabemusters) unter einer Regel, einem Begriff, geschieht 'holistisch' in dem Sinn, daß das gesamte Netzwerk seinen Beitrag für die Aktivierung eines bestimmten Ausgangssignals leistet.

b) Das gesamte Netzwerk verändert sich mit jeder neuen Erfahrung (Eingabemuster) an vielen Stellen und nimmt dadurch die jeweilige Erfahrung in sich auf. Dies schlägt sich dann in einer potentiell veränderten Subsumption einer späteren Erfahrung nieder.

Diese 'äußerlichen' Parallelen sieht Dreyfus [DD87, DD88] als Grund an, im konnektionistischen Paradigma, in subsymbolischen Verarbeitungsmodellen, nach einer adäquaten Modellierung menschlicher Intelligenz zu suchen.[19]
Vom Standpunkt der algorithmischen Informationstheorie ist dies allerdings ein äußerst fataler Fehlschluß ! Dreyfus schließt dabei von der Erkenntnis, daß Intelligenz auf einer sehr komplexen Struktur beruht, darauf, daß man lediglich einen Mechanismus benötigt, der ebenfalls diese Eigenschaft besitzt: namentlich, kompliziert und unübersichtlich zu sein.
Daß das eigentliche Problem nicht darin liegt, ein kompliziertes System zu entwickeln, sondern ein adäquates System, welches als akzidentielle Eigenschaft kompliziert sein muß, scheint er völlig zu übersehen ! Abgesehen davon, *scheinen* künstliche konnektionistische Systeme nur ein komplexeres Ein-/Ausgabeverhalten zu haben - im Sinne der algorithmischen Informationstheorie, wie in Hoffmann [Hof90b] gezeigt wurde. Weiterhin ist festzuhalten, daß man natürlich keine spezielle konnektionistische Architektur oder Ähnliches benötigt, um ein komplexes System zu entwickeln. Der konnektionistische Ansatz muß insofern erst noch nachweisen, daß er wenigstens gleichermaßen erfolgversprechend ist, wie die traditionellen Ansätze der symbolischen künstlichen Intelligenz. Denn abgesehen von dem bei beiden Ansätzen gleichermaßen vorhandenen Problem der Entwicklung eines *adäquaten* komplexen Systems, welches notwendigerweise einen enormen Entwicklungsaufwand erfordert, kommt bei dem konnektionistischen Ansatz das folgende erschwerend hinzu: Das letztlich resultierende Verhalten eines konnektionistischen Systems scheint erheblich schwerer zu übersehen zu sein, als das Verhalten eines symbolischen Systems. Dadurch bliebe jedoch kein Vorteil des Konnektionismus gegenüber dem symbolischen Ansatz mehr übrig !

biologischen Strukturen Redundanz im Sinne der Beschreibungskomplexität aufweisen.

[18] Siehe z.B. Rumelhart et al. [RMt86].

[19] Siehe Dreyfus [DD88] oder [DD87].

8.5 Kognitive Selbstorganisation

Maturana [Mat70] und Varela untersuchten in [MV82, MV80, MV87] die *Selbstorganisation*[20] biologischer Systeme.[21]
Sie übertrugen ihre biologischen Untersuchungsergebnisse in eine Theorie der Erkenntnis biologischer Systeme im allgemeinen. Biologische Systeme - insbesondere das menschliche Gehirn - beruhen auf einer sehr großen Anzahl von kleinen, langsamen Verarbeitungseinheiten - den Neuronen. Diese Verarbeitungseinheiten sind hochgradig miteinander vernetzt.[22]
Jedes der einzelnen Verarbeitungseinheiten arbeitet nur lokal, d.h. seine Funktionsweise wird nur durch seine unmittelbare Umgebung, seine angeschlossenen Nachbareinheiten beeinflußt. Durch das gleichzeitige parallele Wirken aller im Gesamtnetz vorhandenen Neuronen entsteht ein von dem lokalen Verarbeitungsmechanismus 'abgekoppeltes' Gesamtverhalten. Das Gesamtverhalten kann nicht durch die Betrachtung nur lokaler Prozesse erklärt werden. Dies wird auch als *Emergenz* oder *emergentes* Verhalten bezeichnet.[23]
Auch Varela [Var90] stützt sich ähnlich wie Dreyfus [DD88] bei der Argumentation für seine Sichtweise auf phänomenologische Einsichten die auf Heidegger, Merleau-Ponty und andere zurückgehen.
Seine grundlegende Idee ist, daß sich kognitive Systeme nicht auf Repräsentationen beziehen, nicht auf eine äußere 'Realität' referieren, sondern eine interne komplexe kognitive Organisation entwickeln, welche die äußeren Reize strukturiert und interpretiert. Somit *kreieren biologische Systeme ihre eigene Erfahrungswelt.* Es kommt nicht auf eine wie auch immer geartete Korrespondenz interner Strukturen, Repräsentationen oder Aktivierungsmuster[24] mit einer äußeren 'Realität' an. Es genügt vielmehr, wenn das emergente Verhalten eines kognitiven Systems hinreichend an die Umwelt angepaßt ist. Diese Angepaßtheit wird auch als *Handlungswirksamkeit* oder *viable Struktur* bezeichnet.[25] Varela nennt als Beispiel dieser Koppelungsprozesse, die also nicht auf eine 'realistische' Repräsentation einer 'äußeren Welt' abzielen, das menschliche Farbensehen:[26] Physiologische Untersuchungen zeigten, daß Menschen beispielsweise ein graues Blatt Papier, wenn es vor einen roten Hintergrund gelegt wird, nicht mehr als grau sondern als grün wahrnehmen.
Er weist auf physiologische Untersuchungen hin, nach denen bestimmte Vögel vier verschiedene Farbrezeptortypen haben - statt wie beim menschlichen Farbwahrnehmungsap-

[20] Eine umfassender Überblick über die historische Entwicklung der Idee der Selbstorganisation ist in [Pas91] gegeben. Serra & Zanarini [SZ90] weisen technische Ansätze zur Selbstorganisation nach.

[21] Siehe auch von Förster [von60, vZ62], Nicolis & Prigogine [NP90] und Eigen [Eig71] für andere naturwissenschaftliche Ansätze zur Selbstorganisation komplexer Strukturen. Frühe naturwissenschaftliche Arbeiten finden sich z.B. in Ashby [Ash47].

[22] Das menschliche Gehirn umfaßt nach Schätzungen rund 100 Milliarden Neuronen. Jedes Neuron ist mit bis zu 10000 anderen Neuronen verbunden. Siehe Feldman et al. [FFGL90] Seite 434 ff.

[23] Siehe beispielsweise Varela [Var90] Seite 80.

[24] Dieser Begriff bezieht sich auf konnektionistische Systeme, in denen der gleichzeitigen Aktivierung bestimmter Neuronen eine bestimmte Bedeutung zugeordnet werden kann.

[25] Varela [Var90] Seite 108.

[26] Varela [Var90] Seite 107.

parat nur drei. Dies würde vermutlich in unterschiedlichen, nicht aufeinander reduzierbaren Farbwahrnehmungswelten resultieren.
Insofern rekonstruieren oder beinhalten mentale Repräsentationen nicht Informationen aus einer externen *Wirklichkeit*. Vielmehr entstehen mentale *Repräsentationen* durch Interaktion mit der Umwelt. Die Interaktionen haben dabei mehr eine *auslösende* als eine *instruktive* Rolle. Es findet kein Informationsaustausch statt, der eine strukturelle Ähnlichkeit zwischen mentalen Strukturen und den Strukturen der Umwelt bedingt. Es handelt sich vielmehr bloß um eine Koppelung zwischen 'Umweltereignissen' und des 'Geistes Erzeugung von neuen Seinsweisen'. Der Geist ist autonom und bestimmt innerhalb eines unendlichen Repertoires von Verhaltensweisen, eine Verhaltensweise, die an die wahrgenommenen Umweltereignisse gekoppelt ist. Koppelung bedeutet dabei nicht Korrespondenz. Es ist keine strukturelle Übereinstimmung zwischen Umweltsignalen und mentalen Strukturen erforderlich. Grob gesagt ist die Bedeutung der mentalen Repräsentationen eines autonomen Systems mehr vom Subjekt abhängig als von der um das Subjekt herum bestehenden 'Realität'.
Varela sieht in seiner Sichtweise der Selbstorganisation kognitiver Systeme auch die Grundlage einer umfassenden künstlichen Intelligenz:

> Will man außerdem eine künstliche Intelligenz schaffen, die Maschinen herstellt, die in dem Sinne intelligent sind, daß sie mit dem Menschen zusammen (so wie die Tiere) eine gemeinsame Welt des Verstehens und Handelns aufbauen, dann sehe ich keinen anderen Weg dafür, als diese Maschinen in gleicher Weise durch einen Prozeß evolutionärer Transformationen zu entwickeln bzw. zu erziehen, wie es die handlungsbezogene Perspektive nahelegt.[27]

Varela geht also davon aus, daß bestimmte Regularitäten innerhalb der durch Sinnesorgane aufnehmbaren Signale im Laufe der phylo- bzw. ontogenetischen Entwicklung eines Individuums gefunden werden. Dementsprechend organisiert sich dann ein kognitives System, das eine einigermaßene Stabilität zwischen den Signalen der Sinnesorgane und den kognitiven Prozessen (also so etwas wie erfüllte Erwartungen) aufweist.
Vom komplexitätstheoretischen Standpunkt aus betrachtet, erscheint diese Vorstellung wie folgt:
Zunächst kann von der verteilten Struktur neuronaler Systeme völlig abstrahiert werden. Nur die Kolmogoroffkomplexität eines kognitiven Systems, das sich aufgrund einer Flut von 'Sinnesdaten' selbst zu einem hochkomplexen System entwickelt, soll betrachtet werden. Dabei stellt sich die zentrale Frage, wie groß ein möglicher Komplexitätszuwachs in einem kognitiven System ist, das mit einer entsprechenden Menge von Sinnesdaten konfrontiert wird:

> Wieviel (algorithmische) Information muß ein System bereits inkorporieren, bevor es in der Lage ist, in einer großen Datenmenge *zweckmäßige* Regelmäßigkeiten zu erkennen und diese in eine zweckmäßige, handlungswirksame, viable Veränderung (oder Erweiterung) der eigenen Struktur einfliessen zu lassen.

[27] Varela [Var90] Seite 121.

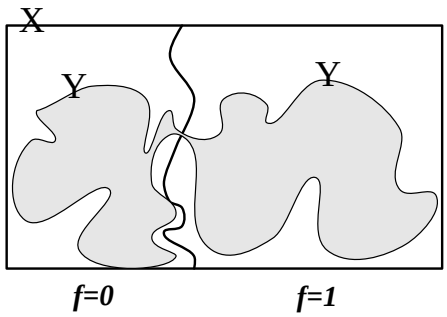


Abbildung 8.6: Die Klassifikationsfunktion f soll aufgrund der Menge Y von einem selbstorganisierenden System bestimmt werden.

Denn in gewisser Weise ähnelt Varelas Vorstellung - in Begriffen der Kolmogoroffkomplexität - dem Münchhausen, der sich an seinem eigenen Schopf aus dem Sumpf ziehen will.
Jedenfalls läßt sich schnell erkennen, daß die Organisation eines kognitiven Systems nicht wesentlich an Komplexität zunehmen kann, wenn die präsentierten Sinnesdaten von einfacher Beschreibbarkeit sind, d.h. wenn sie entsprechend regelmäßig sind.[28]
Ein selbstorganisierendes System braucht in jedem Fall bereits eine gewisse Information, um die Sinnesdaten so strukturieren zu können, daß es sie für eine Weiterentwicklung nutzen kann.
Freivalds und Hoffmann [FH92] stellen einen formalen Rahmen für die Untersuchung von selbstorgansierenden Systemen vor. Erste Ergebnisse über die prinzipiellen Möglichkeiten und Grenzen der Komplexitätssteigerung selbstorganisierender Systeme werden ebenfalls dort präsentiert. Ein einzelnes Ergebnis daraus soll kurz skizziert werden:

Formale Betrachtungen

Um den Leser nicht zu langweilen, werden nicht alle formalen Einzelheiten angegeben. Es soll nur eine Vorstellung von der Art der Untersuchungen vermittelt werden.
Im folgenden wird davon ausgegangen, daß man immerhin von einer konstanten - allerdings unbekannten - Wahrscheinlichkeitsverteilung über alle potentiellen, verschiedenen Sinnesreizungen sprechen kann. (Die unterschiedlichen Sinnesreizungen werden auch formal als *Objekte* bezeichnet.) In Abbildung 8.6 ist in grau eine Menge von Objekten zu sehen, die dem System potentiell präsentiert werden könnten. Das Ziel des selbstorga-

[28] Dies erkennt man, indem man sich vorstellt, daß die einfache Beschreibung in einem kognitiven System bereits in einem besonderen Teil gespeichert ist und reproduziert werden kann. Dann könnte das System sich die Daten selbst präsentieren; völlig losgelöst von jeglichen Umwelteinflüssen. Doch dann kann das System nach Definition der Kolmogoroffkomplexität seine Komplexität oder algorithmische Information nicht steigern - die algorithmische Information erhöht sich ja nicht durch die Ausführung determinierter Berechnungsschritte. Daher kann sich die Komplexität eines kognitiven Systems höchstens um die Kolmogoroffkomplexität der präsentierten Daten erhöhen. Wenn diese aber einfach strukturiert sind - und das müssen sie sein, wenn sie umfangreiche Regelmäßigkeiten aufweisen sollen - so ist ihre Beschreibung von geringer Kolmogoroffkomplexität.

nisierenden Systems ist es, die Klassifikationsfunktion f selbst zu bestimmen, oder eine Approximation zu finden, die möglichst viele der im grauen Bereich liegenden Objekte genau nach f klassifiziert. Das selbstorganisierende System muß also nun aufgrund der nach einer unbekannten Wahrscheinlichkeitsverteilung zufällig vorkommenden Sinnesreizungen, eine 'sinnvolle'[29] Klassifikationsfunktion f bestimmen, die möglichst von höherer Komplexität ist, als das System selbst. Nur durch einen solchen Schritt kann sich ein System durch Selbstorganisation zu einem komplexeren System entwickeln.

Das folgende Ergebnis sagt etwas darüber aus, inwiefern einfache selbstorganisierende Systeme in der Lage sind, bestimmte Klassifikationsfunktionen hoher Komplexität aus zufällig gezogenen Stichproben von unklassifizierten Objekten zu gewinnen oder wenigstens näherungsweise zu gewinnen. Der Leser, der sich nicht für den Formalismus interessiert, möge gleich hinter Theorem 4 springen.

Sei X eine unendliche Menge von Objekten und $X_n \subset X$ diejenige Teilmenge von X, die die ersten n Objekte enthält. X_n sei die Menge von Objekten, die von dem sich selbstorganisierenden System S zu klassifizieren sind.

Definition 4 *Gegeben sei für ein beliebiges n eine feste Teilmenge $Y_n \subseteq X_n$ und eine feste Wahrscheinlichkeitsverteilung P_n über Y_n. Dann sagen wir, daß eine Funktion f eine Zielfunktion f_t* **ε-approximiert**, *wenn gilt:*

$$\sum_{x \in \{x | x \in Y_n \wedge (f(x) \neq f_t(x))\}} P_n(x) \leq \varepsilon$$

Eine ε-Approximation einer Klassifikationsfunktion f_t klassifiziert ein zufällig ausgesuchtes Objekt x höchstens mit der Wahrscheinlichkeit ε anders als f_t.
Wenn eine wichtige Klassifikationsfunktion für lebenserhaltende Klassifikationen der Umweltsituationen nicht hinreichend gut durch das selbstorganisierende System approximiert wird, so ist die Wahrscheinlichkeit entsprechend groß, in Lebensgefahr zu geraten.

Das folgende Theorem gibt eine Obergrenze für die Komplexität von approximierbaren Klassifikationsfunktionen in selbstorganisierenden Systemen an. $K_{Y_n}(f_t)$ ist dabei grob gesagt[30] die Kolmogoroffkomplexität der zu bestimmenden Zielklassifikationsfunktion.

Theorem 4 *Für alle $0 < \varepsilon < \frac{1}{2}$ und $0 < \delta < 1$, eine beliebige aber feste Stichprobengröße s und einer beliebigen Zielfunktion f_t mit einer Komplexität von*

$$K_{Y_n}(f_t) > K(S) + 3 \log_2 n + const,$$

gilt das Folgende:

[29] Wenn man teilweise lebensgefährdende Umstände bei bestimmten Sinnesreizungen annimmt, so könnte man auch statt von 'sinnvoll' von *viabel* oder *lebenserhaltend* sprechen.

[30] Auf die weiteren technischen Details soll hier nicht eingegangen werden.

Es gibt kein selbstorganisierendes System S, das für jede beliebige Wahrscheinlichkeitsverteilung P_n über X_n eine Funktion f aus einer zufällig nach P_n gezogenen Stichprobe der Größe s bestimmt, so daß f die Zielfunktion f_t mit einer Wahrscheinlichkeit von mindestens $1 - \delta$ ε-approximiert.

Beweis: Siehe Freivalds & Hoffmann [FH92].

Das Theorem deutet an, daß es bestimmte Grenzen einer zuverlässigen Selbstorganisation gibt: daß ein System ohne hinreichende Vorinformation, nicht in der Lage ist, durch reine Beobachtung unklassifizierter Objekte zu einer zweckorientierten wesentlich komplexeren Struktur zu gelangen. Es beschränkt die Komplexität der approximierbaren Klassifikationsfunktionen im wesentlichen auf die Komplexität des Systems selbst ! Nur ein geringer Teil ($const + 3 \log_2 n$) geht über die bereits vorher vorhandene Systemkomplexität $K(S)$ hinaus.
Die Voraussetzungen, die in das obige Theorem eingehen, sind natürlich nicht immer erfüllt. Einerseits ist gefordert, daß eine konstante Wahrscheinlichkeitsverteilung über den Objekten besteht - sie könnte mit der Zeit Veränderungen unterworfen sein. Andererseits geht der Beweis des Satzes auf eine Betrachtung des schlechtesten Falls zurück, hier auf die jeweils ungünstigste Wahrscheinlichkeitsverteilung über die für eine gute Approximation wesentlichen Objekten. Es ist somit nicht ausgeschlossen, daß ein selbstorganisierendes System in der Praxis bessere Klassifikationsleistungen erbringt. Ein weiterer Einwand mag die Annahme betreffen, daß es nur *eine* Zielfunktion gibt, die approximiert werden soll. Es ist durchaus möglich, daß es eine ganze Reihe von unterschiedlichen aber gleichermaßen *viablen* Klassifikationsmöglichkeiten gibt, die wahlweise approximiert werden können sollten.

8.6 Kreativität und Komplexität

Im allgemeinen versteht man unter *Kreativität* die Schöpfung von etwas Neuem. Dabei läßt sich zwischen einer *kombinatorischen* und einer *topischen* Kreativität unterscheiden.[31]
Die kombinatorische Kreativität bezieht sich dabei auf den Akt, aus einer vorgegebenen Menge von kombinatorischen Möglichkeiten eine bestimmte Variante oder Variantenklasse herauszugreifen, die für einen gegebenen Zweck geeignet ist.[32]
Im Gegensatz dazu steht die topische Kreativität. Bei der topischen Kreativität wird ein neuer Bezug zwischen Bekanntem gestiftet. Der neue Bezug ist dabei nicht einfach eine Kombination aus bereits vorgegebenen Strukturen. Der neu gestiftete Bezug läßt sich nicht aus Gegebenem durch bestimmte (Kombinations-) Regeln ableiten. Hingegen erlaubt er eine völlig neue Sichtweise auf Gegebenes. Neue Aspekte treten hervor, bisher dagewesene Aspekte verschwinden. Die Stiftung eröffnet ein neues Spektrum von Kombinationen, die jenseits der zuvor gegebenen Kombinationsmöglichkeiten lagen. (Dies geht

[31] Vergleiche auch E. Jelden [Jel90] S. 54–70 und 107–118.
[32] Siehe hierzu beispielsweise Luchins [Luc65].

in der Regel einher mit einer neuen - nicht auf die bisherige Beschreibung reduzierbare - Beschreibung des Gegebenen.) So schreibt beispielsweise Ghiselin aus der psychologischen Kreativitätsforschung:

> *... measure of creative product be the extent to which it restructures our universe of understanding.*[33]

Als Beispiele seien hierfür die folgenden genannt:

- Die Entwicklung mathematischer Begriffe.[34]
- Die Einsteinsche Relativitätstheorie.
- Der Einsatz von - bereits bekannten - Lasern für die ebenfalls bereits bekannte digitale Nachrichtenübertragung.[35]
- Der Einsatz der allgemein verbreiteten Schaufelräder durch Umkehrung des Energieflusses als Turbine.[36]

Insofern ist die topische Kreativität stets bezogen auf eine bisher vorherrschende Sichtweise der Gegebenheiten, die den neu zu stiftenden Bezug nicht in ihrer kombinatorischen Hülle[37] enthält. Die Dichotomie der kombinatorischen versus topischen Kreativität ist somit eine nur relativ zu Bestimmende, in Abhängigkeit gegebener Repräsentationen der Gegebenheiten und der daraus resultierenden Kombinationsregeln. Bei der Betrachtung menschlicher Kreativität kann man sich dabei auf die in der im Indiviuum oder auch in einer Gesellschaft vorherrschenden, bewußten Sichtweise der Gegebenheiten beziehen. Wenn von Kreativität gesprochen wird, so ist im allgemeinen die topische Kreativität gemeint.[38]
Bezieht man nun diese Dichotomie auf eine künstliche Intelligenz, so läßt sich allerdings der genannte Bezug zu einer vorherrschenden und vor allem *bewußten* Repräsentation nicht mehr finden - will man Maschinen nicht Bewußtsein zusprechen.
Solange man sich auf die in Abschnitt 5.2 genannten Annahmen bezieht, nämlich daß sich Intelligenzleistungen und damit auch Kreativität immer in schriftlicher Kommunikation manifestieren kann, so ist jede mögliche kreative Leistung bei einer entsprechenden

[33] In Taylor [Tay64] S. 6.

[34] Beispielsweise erlaubt der Begriff der Kolmogoroffkomplexität auf vielen Gebieten eine neue Sichtweise der Gegebenheiten.

[35] Entnommen aus E. Jelden [Jel90] S. 70.

[36] Entnommen aus E. Jelden [Jel90] S. 70.

[37] Unter kombinatorischer Hülle soll die Gesamtheit aller kombinatorischen Möglichkeiten aufgrund gegebener Kombinationsregeln und Kombinationselemente verstanden werden.

[38] Briskman [Bri82] beispielsweise sieht im kreativen Akt immer eine Transzendenz des bisherigen Rahmens, in dem sich der Kreative befindet. Popper formuliert dies im Kontext wissenschaftlichen Denkens wie folgt: "I do admit that at any moment we are prisoners caught in the framework of our theories; our expectations; our past experiences; our language. But we are prisoners in a Pickwickian sense: if we try, we can break out of our framework at any time. Admittedly, we shall find ourselves again in a framework, but it will be a better and roomier one; and we can at any moment break out of it again." aus K. Popper [Pop70] Seite 56. Rogers [Rog59] sieht Kreativität als nur metaphorisch auffaßbaren Begriff.

Systemstruktur *möglich* und weiterhin eine *kombinatorische* Kreativität. Schließlich läßt sich jede schriftliche Dokumentation einer kreativen Leistung, wie wissenschaftliche Arbeiten, Patentschriften oder literarische Kunst durch geeignete Aneinanderreihung von Buchstaben - also durch Kombination - erzeugen.

In der psychologischen Kreativitätsforschung wird Kreativität oft mehr oder weniger ausdrücklich als mindestens sechsstellige Relation angesehen: [39]

$$K(H, I, R, P, B, S)$$

Dies liest sich wie folgt: Eine Handlung H, die von dem Individuum I ausgehend von einem Rahmen von Wissen, Anweisungen und Absichten R zu dem Produkt P führt, wird von einem externen Beurteiler B unter Bezugnahme auf ein System S von Erwartungen und Zwecken als *kreativ* eingestuft.
Betrachtet man nur ein einzelnes KI-System, so ist es selbstverständlich, daß eine topische Kreativität - topisch bezogen auf das jeweilige System - nicht möglich ist. Bezieht man topische Kreativität jedoch auf die äußerliche Beobachtung und Beurteilung des Kreativitätsaktes und letztlich auf die schriftliche Manifestation dessen, so läßt sich keine allgemeine Aussage darüber treffen, ob eine künstliche Intelligenz eine solche Kreativitätsleistung hervorzubringen in der Lage ist. Es läßt sich stets trivialerweise feststellen, daß es Systeme gibt, die die jeweilige schriftliche Manifestation erzeugen. Dabei könnte man natürlich einwenden, daß in einem solchen Fall der kreative Akt nicht durch das System zustandekam, sondern durch den Entwickler oder Programmierer.
Doch läßt sich dies sicher nicht in allen Fällen sinnvoll behaupten: Betrachtet man beispielsweise einen Bereich, in dem die Entwicklung zu den fortgeschrittensten in der künstlichen Intelligenz zählt - das Schachspielen, so kann man wohl kaum sagen, daß der Entwickler bereits alle resultierenden Rechenergebnisse im Geiste vorwegnahm. Andererseits muß man auch im Schachspiel - obwohl es vielleicht auf den ersten Blick durch seine klardefinierten Regeln als schlechtes Beispiel erscheint, von menschlichen Kreativitätsleistungen, auch topischen Kreativitätsleistungen, ausgehen. Dies geht auch aus der umfangreichen Schachliteratur hervor, die oft genug von schöpferischen, kreativen Zügen, Partien und Spielern spricht. Eine topische Kreativität kommt hier in Betracht, wo Züge gewählt werden, deren strategisches Ziel ein ganz neues, ein nicht in der vorliegenden Position als typisch angesehenes ist.
Heutige Schachprogramme sind so kompliziert[40] in ihrem algorithmischen Regelwerk, daß sie weder ein Entwickler noch ein Schachexperte übersehen könnte. In [LHM+91] weist beispielsweise Hsu, einer der maßgeblichen Entwickler des derzeit stärksten Schachprogramms *'Deep Thought'* darauf hin, daß die Veränderung einer Bewertungsfunktion auf Vorschlag von sehr starken Schachspielern statt zur gewünschten Verbesserung, oft zu einer Stagnation und teilweise sogar zu einer Verschlechterung der Spielstärke führte !

[39] Siehe beispielsweise Dentler & Mackler [DM64] oder Royce [Roy98].

[40] Bei Schachprogrammen müßte man bei 'technischen' Betrachtungen von der resourcenbeschränkten Kolmogoroffkomplexität Gebrauch machen, da mit einem einfachen Programm - aber mit astronomischer Rechenzeit - alle Spielvarianten bis zum Spielende durchsimuliert werden könnten.

Aus dem obigen Beispiel geht also hervor, daß

- Maschinen in der Lage sind, eine bei Menschen als 'kreativ' angesehene Leistung hervorzubringen.[41]
- auch die Systementwickler nicht in der Lage sind, die von dem System hervorgebrachten Leistungen zu antizipieren.

Weiterhin ist zu beobachten, daß für die Dichotomie topische vs. kombinatorische Kreativität ein Bewußtsein geradezu von konstitutiver Bedeutung ist. Ein Bewußtsein von dem aus entschieden werden kann, ob das Ergebnis des Kreativitätsaktes eine Kombination geläufiger Elemente nach geläufigen Regeln ist oder ob ein neuer Bezug gestiftet wird, der nicht bekannt war, aber doch begriffen[42] werden kann, mithin latent bereits im Geiste vorlag.

Aufgrund dieses Sachverhaltes halte ich die Anwendung des Kreativitätsbegriffs (d. h. der Dichotomie topische vs. kombinatorische Kreativität) auf Maschinen für einen Kategorienfehler !

Bei Maschinen ist jede Kreativitätsleistung per definitionem eine kombinatorische. Da jedoch Maschinen offensichtlich bei hinreichender Porgrammkomplexität zu Leistungen

[41] Bei Spielen lassen sich *konstitutive* und *regulative* Regeln unterscheiden. Die Einhaltung der konstitutiven Regeln ist erforderlich, um am Spiel überhaupt teilzunehmen. Die regulativen Regeln hingegen betreffen nur die Spielstrategie.

Somit könnte man bei dem betrachteten Beispiel des Schachprogramms dem genannten Kreativitätsanspruch das folgende entgegenhalten:

Die 'kreative' Leistung der Maschinen ist nur *innerhalb* des durch die konstitutiven Regeln gegebenem Rahmen möglich.

Jedoch ist die scharfe Trennung zwischen konstitutiven und regulativen Regeln zunächst nur aus der Anwendungsperspektive, also von dem jeweiligen Interpretationsbereich des Programms aus möglich. Schließlich läßt sich auch ein Programm entwickeln, das gelegentlich die konstitutiven Regeln des Schachspiels verletzt, oder - wenn man das Ziel zu gewinnen mit zu den konstitutiven Regeln zählt, das gelegentlich 'absichtlich' verliert, um dem Gegner nicht allen Mut zu nehmen.

Somit läßt sich also der Rahmen der konstitutiven Regeln desjenigen 'Spiels', das das jeweilige Programm spielt, immer erweitern, indem eine zusätzliche Regel programmiert wird, die dafür sorgt, daß die bisher konstitutiven Regeln gelegentlich verletzt werden.

Dieser Rahmen besteht aus einer endlichen Zahl von Regeln - aus einem endlichen Algorithmus - der Rahmen aber kann immer weiter gespannt werden.

Es zeigt sich auch hier, daß ohne Einschränkung des Algorithmenbegriffs mittels des Begriffs der Kolmogoroffkomplexität keine Grenze von möglichen algorithmischen 'Kreativitätsakten' gezogen werden kann. Damit verliert aber auch die Frage nach der Möglichkeit 'algorithmischer Kreativität' ihren Sinn.

[42] So schreibt beispielsweise Briskman [Bri82] Seite 136: "One of the most striking, and in any way, paradoxical, features of great creative advances is how often they appear to be, with hindsight, almost obvious." Albert Einstein drückt dies so aus: "In the light of knowledge attained, the happy achievement seems almost a matter of course, and any intelligent student can grasp it without too much trouble. But the years of anxious searching in the dark, with their intense longing, their alternations of confidence and exhaustion, and the final emergence into the light - only those who have themselves experienced it can understood that." zitiert nach Hoffmann & Dukas [HD72] Seite 124.

im Stande sind, die mit topischen menschlichen Kreativitätsleistungen verglichen werden können, erscheint das einseitige Absprechen der Fähigkeit der Maschine zu topischer Kreativität wenig zweckmäßig.
Maschinen gelangen sicherlich auf eine Weise zu ihrem Ergebnis, die bestenfalls metaphorisch mit dem menschlichen Kreativitätsakt verglichen werden kann.

Bei Menschen hingegen gibt es durchaus für beide Kategorien, für die topische wie für die kombinatorische Kreativität klare Beispiele, obwohl auch bei Menschen eine Grauzone besteht.
Insofern macht die Unterscheidung beim Menschen noch Sinn - hier läßt sich teilweise introspektiv ein klarer Unterschied wahrnehmen. Bei Maschinen hingegen läßt sich aufgrund der unvergleichbaren Vorgehensweise keine derartige Unterscheidung ausmachen, obgleich mögliche Rechenergebnisse von einem äußerlichen Beobachter in eine der beiden Kategorien eingeordnet werden können.
Beim Menschen ist durch die Beschränkung des Bewußtseins[43] auch die Menge der noch geläufigen Regeln und Gegenstände beschränkt; bei der Maschine ist dies hingegen von der Konzeption her unbeschränkt.[44] Daher erscheint mir der Versuch einer Analogiebildung von menschlichen Kreativitätsakten und maschineller Vorgehensweise verfehlt. Hier sei auch noch darauf hingewiesen, daß Computer beim Schachspiel schon sinnvolle Strategien entwickelt haben, die anscheinend zu komplex sind, um vom menschlichen Geist begriffen werden zu können.[45] Dies wäre eine ganz neue Art von Kreativität, die durch den Computer möglich werden könnte !

Vielleicht hätte es noch am ehesten Sinn, die maschinelle Fähigkeit zu Leistungen, die von Beobachtern als topische Kreativität eingestuft werden könnten, zumindest unter anderen Faktoren auch durch die dem System immanente Kolmogoroffkomplexität zu bewerten. Immerhin wäre dies mit ein Maß dafür, inwiefern man den Systementwicklern noch einen Überblick über mögliches Systemverhalten zusprechen kann.
Sicherlich hängt es auch von der Systemstruktur - z.B. wie allgemein bestimmte Regeln sind etc. - und nicht nur von der Systemkomplexität ab, ob ein Computer zu Leistun-

[43] 1945 führte J. Hadamard eine Studie über die Arbeitsmethoden herausragender Wissenschaftler in den USA durch. Unter Ihnen war auch Einstein, der schrieb: "The words of the language as they are written or spoken do not seem to play any role in my mechanism of thought, which relies on more or less clear images of a visual and some of muscular type. It seems to be that what you call full consciousness is a limiting case which can never be fully accomplished because consciousness is a narrow thing." zitiert nach Koestler [Koe82] Seite 13.

[44] Bei einer physikalisch realisierten Maschine liegen natürlich Grenzen für die Zahl der speicherbaren Regeln und Symbole vor, aber diese Grenze ist in den letzten Jahrzehnten außerordentlich gestiegen und wird voraussichtlich auch in näherer Zukunft noch ähnliche Steigerungsraten zu verzeichnen haben.

[45] Das Endspiel Dame und König gegen Turm und König wurde bisher übereinstimmend von allen Schachmeistern für leicht gewinnbar für die Damenpartei gehalten. Jedoch haben Schachprogramme für die Turmpartei Strategien gezeigt, die es selbst internationalen Schachmeistern nicht erlaubte, die Endspiele zu gewinnen. Die Strategie ist so kompliziert, daß bisher kein Mensch darauf gekommen war. Auch nachdem sie von Computern vorgeführt wurde, ist sie für menschliche Schachspieler nicht nachvollziehbar ! Siehe hierzu Nunn [Nun91].

gen imstande ist, die selbst seine Entwickler als *topische* Kreativitätsakte kategorisieren würden.

Kapitel 9

Die Grenzen der künstlichen Intelligenz

Die Frage nach den Grenzen einer künstlichen Intelligenz wird häufig kontrovers diskutiert. Vor dem Hintergrund der vorangegangenen Überlegungen und dem Begriffs der Kolmogoroffkomplexität sollen die Gründe für die kontroversen Standpunkte erhellt werden. Es wird deutlich werden, daß unterschiedliche implizite Voraussetzungen dazu führen, Grenzen der künstlichen Intelligenz zu sehen, oder auch nicht zu sehen.
Im Abschnitt 9.1 wird zunächst eine Kategorisierung der unterschiedlich formulierten Fragen nach den Grenzen künstlicher Intelligenz vorgenommen werden. Abschnitt 9.2 stellt eine scharfe Präzisierung der Fragestellung vor. Im dritten und vierten Abschnitt wird die Adäquatheit des der Arbeit zugrundeliegenden Turingmaschinenmodells erörtert. Abschnitt 9.5 behandelt schließlich Argumente gegen die Möglichkeit einer KI, die sich auf bestimmte behauptete Eigenschaften des menschlichen Bewußtseins stützen.

9.1 Die Frage nach den Grenzen der künstlichen Intelligenz

Bereits vor der eigentlichen Geburtsstunde der künstlichen Intelligenz 1956 fragte A. M. Turing nach ihren Grenzen ! In seinem Artikel *Computing machinery and intelligence* [Tur50] diskutierte Turing verschiedene Einwände gegen die Behauptung, daß Maschinen denken können. Schlußendlich kam Turing zu der Ansicht, daß es keinen Grund gibt, Maschinen *Denkfähigkeit* abzusprechen, wenn sie nur einen bestimmten Test, bei dem die Maschine menschliches Verhalten imitieren soll, bestehen. Dieser Test ist heute auch als *Turingtest* [1]

[1] Siehe u.a. Moor [Moo87] für eine Diskussion um den Turingtest. Eine philosphisch-soziologische Diskussion um den Turingtest findet sich Wiener [Wie84]. French [Fre90] argumentiert, daß der Turingtest nicht Intelligenz im allgemeinen prüft, sondern nur eine sehr eingeschränkte, kulturspezifische Intelligenz. Harnad [Har91] schlägt eine Erweiterung des Turingtests auch auf motorische Fertigkeiten vor, um das 'other mind' Problem einer pragmatischen Lösung zuzuführen.

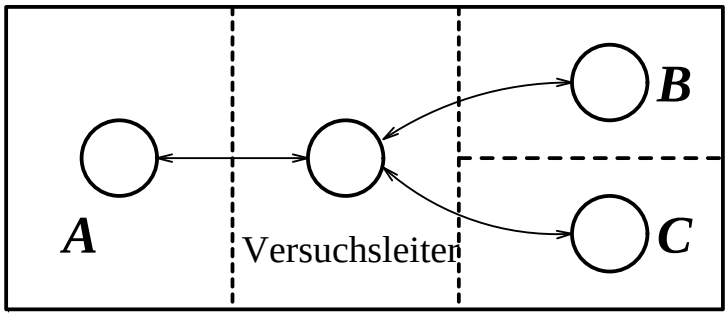


Abbildung 9.1: Szenario des Turingtests. Person A soll herausfinden, ob sich hinter B oder hinter C die Maschine verbirgt.

bekannt.
Dabei wird eine Person A in einem abgeschirmten Raum gebeten, über einen Fernschreiber Fragen an jeweils eines von zwei Systemen B bzw. C zu richten. Hinter einem der Buchstaben versteckt sich ein Mensch und hinter dem anderen eine Maschine, die auf die jeweils gestellten Fragen antworten werden. (Siehe Abbildung 9.1.) Die Aufgabe von A ist nun, die Maschine richtig zu identifizieren. Zur Entscheidungsunterstützung darf A beliebige Äußerungen an wahlweise eines der Systeme B oder C richten, um aufgrund der Reaktionen Rückschlüße auf die Identität von B bzw. C zu ziehen. Ist A nicht in der Lage in mehr als 50% der Fälle richtig zu entscheiden, so hat die Maschine den Test bestanden und man kann der Maschine die Denkfähigkeit nicht absprechen - so Turing.

Die Frage nach den Grenzen der künstlichen Intelligenz wird allerdings keineswegs einheitlich - z.B. in der Form des Turingtests - gestellt. Die meisten Fragetypen lassen sich wie folgt paraphrasieren:

- Können Maschinen denken ?[2]
- Ist das menschliche Denken algorithmisch/algorithmisierbar ?[3]
- Gibt es einen (wesentlichen)[4] Unterschied zwischen Mensch (oder menschlichem Geist) und Maschine ?[5]

Welche dieser Formulierungen man auch betrachtet, es kommt - neben eventuell zusätzlichen Fragen[6] - immer wieder die Frage zur Diskussion, ob eine *algorithmische Beschreibung*

[2] Siehe beispielsweise Turing [Tur50].

[3] Diese Formulierung findet sich unter anderem bei V. Cherniavsky [Che78], H. L. Dreyfus [DD87] und bei R. Penrose [Pen89].

[4] Solange wir zwischen Mensch und Maschine unterscheiden können, gibt es natürlich immer einen Unterschied.

[5] Unter anderen versuchen Jacquette [Jac87], Lucas [Luc61] Penrose [Pen89] und Searle [Sea84] einen Unterschied aufzuzeigen. Siehe hierzu auch Abschnitt 9.5.

[6] Insbesondere Fragen, die Bewußtsein, Verantwortung und dergleichen betreffen.

menschlicher Intelligenzleistungen möglich ist ?[7]
In jedem Fall fällt auf, daß auch diese eingeschränkte Fragestellung immer noch sehr vage ist. Es ist zwar klar, was unter einem Algorithmus, einer Maschine[8] bzw. unter einer algorithmischen Beschreibung verstanden wird. Jedoch ist zunächst unklar, was menschliche Intelligenzleistungen sein sollen. Hier sind gleich mehrere häufig nicht weiter erläuterte Probleme versteckt.

- Welche menschlichen Intelligenzleistungen sind denn gemeint ?
  - Die Leistung eines durchschnittlichen Menschen auf einem Spezialgebiet ?
  - Die Leistung eines Experten auf seinem Spezialgebiet ?
  - Die Leistung des jeweils besten Experten auf einem Gebiet ?
  - Oder eine imaginäre Intelligenzleistung, die von keinem existierenden Menschen tatsächlich geleistet wird. Eine Leistung, die man nur aufgrund irgendwelcher nicht weiter diskutierter Generalisierungen der menschlichen Intelligenz im allgemeinen zuschreiben möchte ?

  Wenn man Schachspielen als Beispiel betrachtet, so kann man festhalten, daß bereits vor einigen Jahren, Schachprogramme entwickelt worden waren, die der Leistung eines durchschnittlichen Menschen deutlich überlegen waren [Lev88b]. Einige Jahre später waren auch die meisten Experten (Meisterspieler) von Schachprogrammen übertroffen worden [MS90]. Heute wird an Schachprogrammen gearbeitet, die den Weltmeister besiegen sollen [HACN90]. Zur Zeit können nur noch ein paar hundert Spieler in der Welt mit dem derzeit stärksten Schachprogramm einigermaßen mithalten [HACN90].

  Allgemeiner gesprochen, kann man die folgenden Fragetypen unterscheiden:

  a) Gibt es Aufgaben, die von jedem Menschen besser als von jeder Maschine bewältigt werden können ?

  b) Gibt es Aufgaben, die von mindestens einem Menschen besser bewältigt werden, als von jeder Maschine ?

[7] Diese Frage ist einerseits eine metaphysische Frage: A posteriori läßt sich immer ein Algorithmus angeben, der zumindest das Ergebnis der Denkprozesse beschreibt. Im vorhinein hingegen, kann man ohne weitere metaphysische Annahmen nie sagen, daß zukünftige Denkprozesse durch diesen oder jenen Algorithmus beschrieben werden. Andererseits sind hier im allgemeinen keine individuellen Denkprozesse einzelner Individuen gemeint, sondern so etwas wie Idealisierungen dessen. Wenn man aber die Idealisierung dessen operational beschreiben kann, so hat man damit auch gleichzeitig einen Algorithmus.

Ist man hingegen, insbesondere in der künstlichen Intelligenz - im Gegensatz zur Kognitionswissenschaft - daran interessiert, nützliche Systeme zu entwickeln, so ist es schwer denkbar, daß das entsprechende Ein-/Ausgabeverhalten nicht algorithmisch sein könnte. Denn dies würde bedeuten, daß es nicht determiniert sei, was in einer gegebenen Situation angemessen ist; also nicht nur, daß es vielleicht mehrere gleichermaßen wünschenswerte Ein-/Ausgabeverhalten gibt oder daß es von der zeitlichen, räumlichen, sozialen etc. Situation abhängt, welches Verhalten geboten ist, oder das es kein Mensch weiß. Letzteres würde Ingenieursprobleme betreffen, aber nicht die Frage der prinzipiellen Algorithmisierbarkeit.

[8] In der theoretischen Konzeption der Turingmaschine.

c) Gibt es Maschinen, die alle denkbaren Aufgaben besser als zumindest ein Mensch bewältigen ?

d) Gibt es Maschinen, die alle denkbaren Aufgaben besser als jeder Mensch bewältigen können ?

In der Regel wird der Fragetyp d) betrachtet, wenn gegen die Möglichkeit einer menschlichen Intelligenz argumentiert wird. Insgesamt ist allerdings festzuhalten, daß häufig eine Anspruchshaltung an maschinelle Intelligenz vertreten wird, der zumindest die allermeisten Menschen nicht genügen würden. Man denke nur beispielsweise an die alltäglichen Fehldiagnosen in der medizinischen Praxis. Oder die fehlerhaften Gedankengänge bei mathematischen Beweisen oder sonstigen intellektuellen Tätigkeiten.

- Wird nur eine Intelligenzleistung gefordert, die sich auch tatsächlich beobachten läßt, oder wird mehr gefordert ? Im ersten Fall würde man also damit auch den Turingtest akzeptieren müssen. Im zweiten Fall stellt sich die Frage, wie das Mehr aussehen kann.

  - Searle [Sea80, Sea82] hält Intentionalität für einen wesentlichen Bestandteil von Intelligenz.[9]
  - Es bietet sich einerseits an, von allgemeinen Prinzipien zu sprechen: Beispielsweise, daß der menschliche Geist prinzipiell dazu in der Lage ist, die Wahrheit einer beliebigen prädikatenlogischen Formel zu entscheiden - z.B. aufgrund einer intensionalen Einsicht in die Bedeutung der prädikatenlogischen Formeln.[10]

    Hier wird wohl jeder schnell einräumen, daß eine praktische Entscheidung von größeren Formeln zu den Lebzeiten eines Mathematikers nicht möglich ist. Was heißt es aber dann, die Entscheidung könne *prinzipiell* korrekt getroffen werden.
  - Ein anderes Argument beruht auf der Behauptung einer *prinzipiellen* nichtalgorithmischen Komponente im menschlichen Geist, die sich insbesondere im kreativen Akt zeigt. Hier wird behauptet, daß - ganz gleich welchen Algorithmus man zur Beschreibung von Intelligenzleistungen auch angibt - der menschliche Geist immer in der Lage sei, von den algorithmischen Regeln abzuweichen und dadurch von einer rein kombinatorischen Kreativität zu einer topischen Kreativität zu gelangen.[11]

Die genannten Vagheiten lassen einigen Spielraum für Spekulationen. Es sei an dieser Stelle darauf hingewiesen, daß die Annahme, daß menschliches Denken nicht algorithmisch sei, auch impliziert, daß wir Menschen kein Mittel der exakten Beschreibung des Denkens

[9] Vergleiche auch Searle [Sea87] über Intentionalität als wesentliche Eigenschaft des menschlichen Geistes.

[10] Siehe Jacquette [Jac87].

[11] Siehe beispielsweise Dreyfus [DD87].

haben.[12] Mithin wir zumindest bestimmte Charakteristika unserer Denkprozesse immer nur metaphorisch beschreiben können.[13]

9.2 Präzisierung der Frage nach den Grenzen

Argumentationen, die Grenzen für die KI behaupten, berufen sich primär entweder auf Problemklassen für die es erwiesenermaßen keinen allgemeinenen Lösungsalgorithmus gibt, z.B. für das Entscheidungsproblem der Prädikatenlogik erster Stufe.[14]
Oder aber sie behaupten, menschliche Intelligenz würde zumindest gelegentlich von jedem beliebigen Regelschema abweichen; dafür wurden häufig bestimmte kreative Leistungen als Beispiele angegeben.
Die erste Art von Argumenten hat Schwierigkeiten nachzuweisen, daß die Unzulänglichkeiten von Algorithmen nicht gleichermaßen auch auf menschliche Intelligenz zutreffen.
Bei der zweiten Art von Argumenten könnte eine präzisere Fassung wie folgt aussehen:
Für einen gegebenen Algorithmus A oder eine Klasse von Algorithmen von dem/der Intelligenz behauptet wird, läßt sich eine Aufgabe angeben, die von menschlicher Intelligenz besser, bzw. überhaupt erst angemessen, behandelt wird.
Auf der anderen Seite wird für die unbegrenzten Möglichkeiten der KI argumentiert, daß jedes geforderte Verhalten - auch beliebige Ausnahmefälle - von einem entsprechend entworfenen Algorithmus gezeigt werden kann.
Die eine Seite sucht zu einem gegebenen Algorithmus eine Intelligenzleistung, die er nicht leisten kann, während die andere Seite zu einer gegebenen Intelligenzleistung einen Algorithmus angibt.
Solange man die Klasse von zu betrachtenden Algorithmen nicht näher einschränkt, sondern beliebige Algorithmen zuläßt, scheint die Diskussion ziemlich fruchtlos, da die Frage unpräzise gestellt ist.
Durch den Begriff der Kolmogoroffkomplexität zur Differenzierung des Diskussionsgegenstandes läßt sich die fruchtlose Streitfrage zu eine sinnvolle Frage wie folgt reformulieren.[15]

> Kann eine spezifische Klasse von Aufgaben A durch ein Programm P gelöst werden, das höchstens von der Länge l ist ?[16]

Erst nach dieser Präzisierung erhält eine solche Fragestellung - zumindest solange sie auf *endliche* Strukturen bezogen ist - wirklichen Sinn. Ob der menschliche Geist 'wirklich' *unendliche* Strukturen von *unendlicher* Kolmogoroffkomplexität hervorzubringen vermag, muß wohl immer in den Bereich der Spekulationen verwiesen werden. Endliche Strukturen - zum Beispiel die symbolische Beschreibung endlich vieler kreativer Akte - sind

[12] Vergleiche Kapitel 3.
[13] Vergleiche dazu, daß z.B. C. R. Rogers [Rog59] den Terminus 'Kreativität' generell nur für metaphorisch zugänglich hält.
[14] Siehe hierzu Abschnitt 9.5.
[15] Die Grundidee wurde in Hoffmann [Hof92] vorgestellt.
[16] Dies läuft auf die Bestimmung der Kolmogoroffkomplexität einer Lösung der jeweils gestellten Aufgabe hinaus.

jedenfalls immer von endlicher Kolmogoroffkomplexität und damit auch algorithmisch beschreibbar. Und solche endlichen Strukturen scheinen die einzigen Strukturen zu sein, deren menschlichen Ursprung man empirisch belegen kann bzw. jemals belegen können wird. Daher begnügte Turing sich auch mit seinem Testkriterium.

In der Praxis sind ohenhin nicht nur Systeme interessant und wertvoll, die die Intelligenzleistung eines beliebigen Menschen in jeder Hinsicht übertreffen oder zumindest ebenbürtig sind. Vielmehr genügt es häufig, wenn intelligente Systeme den Leistungen von vielen Menschen ebenbürtig sind, oder lediglich menschlichen Intelligenzleistungen unterstützen und zu diesem Zweck Routineaufgaben übernehmen.

9.3 Die Grenzen des Turingmaschinenmodells

Bei der Diskussion um die Grenzen der KI wird gelegentlich die schlichte Inadäquatheit der Turingmaschine für die Simulierung kognitiver Prozesse behauptet.[17] Im einzelnen werden unter anderem die folgenden Einschränkungen des Turingmaschinenmodells hervorgehoben:

- Es wird ein *endlicher* Zeichenvorrat benutzt.
- Die Turingmaschine hat 'nur' eine *endliche* Menge von internen Zuständen.
- Die Eingaben werden durch endliche Zeichenketten beschrieben; es sind also keine analogen Eingabewerte möglich.
- Die Ausgaben werden durch endliche oder unendliche *diskrete* Zeichenketten beschrieben.
- Die Turingmaschine arbeitet in *diskreten* Zeitschritten. Es sind keine asynchron auftretenden Eingangssignale möglich, die eventuell zu einem 'nicht-deterministischen' Verhalten führen könnten.
- Die Turingmaschine arbeitet *deterministisch*, abhängig von den oben genannten Größen.

Für jeden der genannten Punkte läßt sich ein Szenario vorstellen, bei dem ein physikalisches und/oder ein biologisches System zumindest auf den ersten Blick nicht den jeweils genannten Einschränkungen unterliegt und damit nicht durch eine Turingmaschine adäquat modelliert werden kann.[18] Im wesentlichen basieren in der Literatur vorgeschlagene Szenarien auf *analogen* Ein- und Ausgangssignalen sowie analogen internen Zuständen von Berechnungseinheiten - wie etwa Neuronen.[19] Mithin wird argumentiert,[20] daß die

[17] Beispielsweise in Smolensky [Smo88].

[18] Ob sich solche Systeme adäquat simulieren lassen, hängt letztlich unter anderem von der Adäquatheit der Quantenphysik ab.

[19] Siehe beispielsweise Mead [Mea89].

[20] Siehe z.B. Brooks [Bro91].

Intelligenz gerade durch diese Andersartigkeit des physikalischen bzw. biologischen Systems zustande kommt, bzw. daß dies der notwendige Entwicklungsboden für Intelligenz innerhalb eines kognitiven, selbstorganisierenden Systems ist.
In der Tat muß eingeräumt werden, daß die genannten Einwände nicht unmittelbar entkräftet werden können, und daß die in der vorliegenden Arbeit angestellten Überlegungen zunächst nur unter der Prämisse der Adäquatheit des Turingmaschinenmodells gelten.
Allerdings erscheinen Argumente durchaus zweifelhaft, die sich darauf zurückziehen, daß Intelligenz in dem Unterschied zwischen diskreten und analogen Eingangssignalen begründet sein soll. Schließlich lassen sich analoge Signale immer beliebig genau durch diskrete Signale approximieren.[21] Nun basiert die Behauptung darauf, daß dieser Unterschied einen 'Quantensprung' vom algorithmischen Regelfolgen zu menschlicher Intelligenz ausmacht. Daß dadurch ein anderes, ein *nicht-algorithmisches* Verhalten möglich wird, ist nicht widerlegbar.[22] Jedoch bleibt es zweifelhaft, daß dadurch gerade eine erforderliche *Verbesserung* des Verhaltens, also eine 'Intelligenzsteigerung' bewirkt wird !
Noch zweifelhafter wirkt das zweite Argument, daß Intelligenz in der möglichen Asynchronität von Eingangssignalen begründet sein soll: Dadurch läßt sich das Systemverhalten zunächst einmal nicht vorhersagen. Daß dadurch ein System allerdings intelligenter wird, ist schwer einzusehen.
Wie dem auch sei, im nächsten Abschnitt wird auch den 'nicht-Turing-Berechnungsmodellen' ein entsprechendes Komplexitätsargument entgegengesetzt.

9.4 Intelligenz nicht-algorithmischer künstlicher Systeme

Abgesehen von dem Fehlen einer streng formalen Behandlung der genannten Einwände lassen sich doch die folgenden Überlegungen anstellen:
Angenommen es gibt tatsächlich bestimmte Systemstrukturen, die sich aufgrund analoger und/oder asynchroner Eingangssignale in einer geeigneten Umgebung durch selbstorganisierende Eigenschaften zu intelligenten Systemen entwickeln können. Seien diese Systeme durch die folgenden Minimalvoraussetzungen charakterisiert, die von den Beschränkungen des Turingmaschinenmodells weitestgehend absehen:

- Es gibt diskrete Ausgabesignale des Systems.

[21] Wenn man an einen heutigen CD-Spieler denkt, so ist es dem menschlichen Hörvermögen doch häufig nicht mehr ohne weiteres möglich, einen Unterschied zu den originalen analogen Signalen zu erkennen. Und die Approximation des Analogsignals durch den CD-Spieler könnte von der theoretischen Konzeption her noch beliebig viel genauer sein. Eine andere Frage ist, ob biologische Systeme oder das menschliche Gehirn auf solche feinen Unterschiede in Signalen überhaupt durch signifikant intelligentere Leistungen reagieren können.

[22] Es gibt irrationale Zahlen deren Beschreibung im Sinne der Kolmogoroffkomplexität einen *unendlichen* Informationsgehalt hat, wie man aufgrund der Tatsache sieht, daß es jeweils nicht komprimierbare Zeichenketten jeder beliebigen Länge gibt. Sie könnten also nur durch ein *unendliches* Programm dargestellt werden, welches nicht mehr unter die Definition eines Algorithmus fällt.

- Zumindest einige Eingangssignale zu dem System sind diskret.
- Es gibt zu jedem denkbaren diskreten Systemverhalten, d.h. auf diskrete Eingangssignale mit diskreten Ausgangssignalen zu reagieren, je verschiedene Systementwürfe, die das entsprechende diskrete Systemverhalten hervorbringen. Wie das jeweilige System auf analoge Eingangssignale reagiert, sei währenddessen nicht näher spezifiziert.

Aus diesen schwachen Voraussetzungen läßt sich bereits ein ganz ähnliches Argument entwickeln, wie es der algorithmischen Informationstheorie zugrunde liegt:
Es ist zunächst festzuhalten, daß die gewünschten Intelligenzleistungen, wie in Abschnitt 5.2 skizziert wurde, von einem solchen nicht-algorithmischen System gleichermaßen abgefordert werden können sollte. D.h. zumindest unter anderem soll das System auf diskrete Eingangssignale mit bestimmten diskreten Ausgangssignalen reagieren.
Mithin läßt sich auch hier festhalten, daß ein Intelligenzverhalten einer bestimmten Kolmogoroffkomplexität gefordert wird. Das Argument der Vertreter der analog/asynchron-Systeme ist nun, daß unter den nicht-algorithmischen Voraussetzungen ein System sich zu einem entsprechend intelligenten System entwickeln kann, was unter algorithmischen Voraussetzungen nicht möglich sei.

Wenn nun die dritte der obigen Bedingungen benutzt wird, so läßt sich festhalten, daß nicht nur dieses eine - intelligente (komplexe) - Verhalten möglich ist, sondern daß auch Systementwürfe möglich sind, die zu anderen Verhaltensweisen führen. Nimmt man jeweils einen spezifischen Systementwurf für jedes mögliche resultierende Verhalten an, so kommt man zu einer 1:1 Abbildung von Systementwurf und später - nach einer 'selbstorganiserenden Intelligenzentwicklung' - zu dem resultierenden Verhalten.
Sei das resultierende intelligente Systemverhalten durch n Binärzeichen beschreibbar. Dann gibt es 2^n verschiedene mögliche Systemverhalten und damit muß es mindestens auch 2^n verschiedene Systementwürfe geben. Dabei korrespondiert dann ein Systementwurf jeweils zu einer möglichen Verhaltensspezifikation. Wenn man nun davon ausgeht, daß ein sich erfolgreich selbstorganisierendes System durch einen einfachen Entwurf beschrieben werden kann, so kommt man zu dem folgenden Schluß:

> Den 2^n verschiedenen Verhaltensmöglichkeiten stehen damit 2^n verschiedene Systembeschreibungen gegenüber. Das heißt, die allermeisten Systembeschreibungen sind notwendigerweise mindestens von der Länge n bei binärer Beschreibung.
>
> Nach Voraussetzung ist aber gerade das *intelligente* Systemverhalten durch ein System von einfacher, d.h. kurzer Systembeschreibung zu erreichen.
>
> Daher müssen die allermeisten längeren Systembeschreibungen zu Systemen führen, die sich nicht zu intelligenten Systemen selbstorganisieren.
>
> Somit wären die einfachen Systeme, diejenigen Systeme, die sich selbst zu komplizierten intelligenten Systemen entwickeln.

> Im Gegensatz dazu müssen sich jedoch die vielen verschiedenen von vornherein komplizierten Systeme selbst zu einfachen oder komplizierten, aber in jedem Fall unintelligenten oder dummen Systemen entwickeln !

Eine solche Situation würde bedeuten, daß die Intelligenz weniger in der Systemstruktur liegt, als mehr in einem 'wundersamen' Analogsignal, das aus irgendeinem physikalischen oder biologischen Element erzeugt wird. Würde man dieses Analogsignal digital darstellen, so erhielte man eine unendliche Ziffernreihe, die alle Intelligenz der Welt in sich trägt. In [LV88] wird eine solche Zahl, aus der beispielsweise alle korrekten Entscheidungen für eine Menge von formal unentscheidbaren prädikatenlogischen Formeln abgeleitet werden können, auch als 'number of wisdom' bezeichnet. Dort wird allerdings nicht über die Intelligenz einzelner biologischer oder physikalischer Entitäten spekuliert. Es sei noch angemerkt, daß es keineswegs genügt, irgendeine Zeichenkette unendlicher algorithmischer Information zu haben; die Dekodierung der 'Weisheit' aus der Zeichenkette muß sich wiederum algorithmisch *einfach* bewerkstelligen lassen.

9.5 Kann ein Bewußtsein jeden Algorithmus transzendieren ?

Die wohl populärste Arbeit aus der jüngeren Vergangenheit zu den Grenzen der künstlichen Intelligenz ist die des Mathematikers R. Penrose [Pen89]. Er führt eine Reihe von Argumenten gegen die Möglichkeiten einer künstlichen Intelligenz an. Er bezieht sich dabei allerdings primär auf die Bewußtseinsfrage, d.h. damit auf die *starke KI-These*, d.h. gegen die Annahme, daß Maschinen - bei entsprechendem Aufbau - Bewußtsein zukommen könnte. Nach einer 'Rundreise' durch Mathematik, Logik und Physik führt Penrose in dem letzten Kapitel seines Buches eine Reihe von Argumenten dafür an, daß

- Intelligenz so etwas wie Bewußtsein erfordert und
- Bewußtsein nicht durch eine künstliche Intelligenz hervorgebracht werden kann.

Die Schlagkraft seiner Argumente erscheint allerdings eher etwas enttäuschend. Hinzu kommt, daß er in einer Passage über den Begriff von Intelligenz die Möglichkeit einzuräumen scheint, daß ein Algorithmus tatsächlich Intelligenz simulieren könnte. Es heißt dort:

> ..., if it turns out that AI people *are* eventually able to simulate intelligence ... In that case the issue of 'intelligence' would not be my real concern here. I am primarily concerned with 'consciousness'.

Einige seiner Argumente sind im folgenden aufgeführt:

- Intelligenz erfordert ein Bewußtsein, denn wäre es überflüssig für unser Verhalten, so hätte die Evolution bei der Entwicklung von intelligenten Systemen auf Bewußtsein

verzichtet, bzw. die intelligenten Lebewesen wären zumindest später dahingehend 'degeneriert', daß sie kein unnötiges Bewußtsein mehr hätten.[23]

- Die Entscheidung, ob ein Algorithmus sinnvoll ist ('the validity of an algorithm')[24] ist nur durch Einsicht zu treffen - nicht durch einen Algorithmus. Daher kann die Evolution nicht algorithmisch sein.[25]

- Das Vokabular mit dem wir über bewußte und unbewußte mentale Prozesse sprechen, legt eine nicht-algorithmische Natur der bewußten mentalen Prozesse nahe.[26]

Dasjenige Argument, das noch die größte Überzeugungskraft zu haben scheint, - auch nach seinen eigenen Worten[27] - ist das folgende, das bereits in der philosophischen Literatur einige Diskussion hervorrief:

Lucas [Luc61] führt die Gödelschen Unvollständigkeitsbeweise von formalen Systemen an, die zumindest die elementare Arithmetik enthalten: Gödel zeigte, daß sich zu jedem formalem System eine wahre arithmetische Aussage konstruieren läßt, die jedoch nicht durch das formale System bewiesen werden kann - und damit auch nicht als 'wahr' von diesem System erkannt werden kann. Das Argument von Lucas ist nun, daß wir Menschen aufgrund unserer Einsichtsfähigkeit hingegen sehr wohl die Wahrheit einer derart konstruierten arithmetischen Aussage erkennen können. Damit wäre nach Lucas gezeigt, daß das menschliche Denken nicht-algorithmisch ist.[28]
Gegenargumente die in der Folge entwickelt wurden, berufen sich beispielsweise darauf, daß faktisch der dem Denken eines Mathematikers zugrunde liegende Algorithmus unbekannt ist und insofern die entsprechende wahre arithmetische Aussage nicht konstruiert und damit auch nicht von dem Mathematiker als wahre Aussage erkannt werden kann.[29]

[23] Siehe Penrose [Pen89] Seite 408/409. Auf Seite 408 heißt es: '... If consciousness serves no selective purpose, why did Nature go to the trouble to evolve *conscious* brains when non-sentient 'automaton' brains like cerebella would seem to have done just as well ?'

[24] Siehe Penrose [Pen89] Seite 414.

[25] Siehe Penrose [Pen89] Seite 414/415. Auf Seite 415 heißt es: 'Moreover, the slightest 'mutation' of an algorithm ... would tend to render it totally useless, and it is hard to see how actual *improvements* in algorithms could ever arise in this random way. ... Perhaps some much more 'robust' way of specifying algorithms could be devised, ... The 'robust' specifications are the *ideas* that underlie the algorithms. But ideas are things that, as far as we know, need conscious minds for their manifestation.'

[26] Siehe Penrose [Pen89] Seite 411: *I think that the kind of terminology that we tend to use, which distinguishes our conscious from our unconscious mental activity, is at least suggestive of a non-algorithmic/algorithmic distinction: ... the action of consciousness ... cannot be described by any algorithm.*

[27] In Penrose [Pen89] Seite 418 heißt es: 'To my thinking, this is as blatant a *reductio ad absurdum* as we can hope to achieve, short of an actual mathematical proof !'

[28] Lucas [Luc61, Luc70], Jacquette [Jac87], Penrose [Pen89] und andere haben darauf basierend argumentiert, daß das menschliche Denken nicht mechanisierbar bzw. algorithmisch beschreibbar ist. Gegen eine solche Argumentation wurden allerdings auch starke Einwände unter anderem von Benacerraf [Ben67], Webb [Web68, Web80] und Slezak [Sle87] erhoben.

[29] Siehe Benacerraf [Ben67].

Penrose nun geht von der Annahme aus, daß das menschliche Denken algorithmisch ist und folgert daraus aufgrund der Tatsache, daß mathematische Begriffe, Beweise und Einsichten kommunizierbar sind, daß es einen allen Mathematikern zugänglichen Algorithmus für die Beurteilung mathematischer Wahrheit ('judging mathematical truth')[30] geben müsse. Dieser wäre dann nur ein Teil der möglicherweise sehr komplizierten Algorithmen die den allgemeinen Denkprozessen eines Menschen zugrunde liegen. Weiter, so Penrose, müsse dieser Algorithmus sehr einfach sein, denn es sei ja gerade die Grundidee mathematischer Beweise, daß jeder Beweisschritt so einfach ist, daß er von jedermann ohne Zweifel nachvollzogen werden kann.[31] Damit müsse sich dieser Algorithmus prinzipiell bestimmen lassen, wobei der obige Einwand, daß der dem faktischen Denken eines Mathematikers zugrunde liegende Algorithmus unbekannt sei, nicht mehr angewendet werden könne.

Aus der Sicht der Kolmogoroffkomplexität ist folgendes dieser Argumentationsweise entgegenzuhalten:
Der Gödelsche Beweis zeigt, daß sich unendlich viele mathematische Theoreme konstruieren lassen, die zu ihrem Beweis je 'individuelle' algorithmische Information benötigen.[32] Damit ist es mit einem Algorithmus - der per definitionem nur einen endlichen algorithmischen Informationsgehalt hat - nicht möglich, *alle* Theoreme zu beweisen. Somit kann es also den von Penrose gesuchten kurzen präzise beschriebenen Algorithmus nicht geben.

Gleichermaßen gilt aber auch für Menschen schon allein aufgrund ihrer beschränkten Lebenszeit, daß sie nicht alle wahren Theoreme als wahr erkennen können. Aber auch schon bei sehr langen Beweisen, ist die intersubjektive Übereinstimmung nur noch schwer zu erreichen. Ohnehin treten bei der mathematischen Beweisführung oft genug Fehler auf, die teilweise auch über lange Zeit unbemerkt bleiben. Dies zeigt beispielsweise die Geschichte des Polyedersatzes,[33] bei der die Definition eines Polyeders immer wieder verändert wurde, weil Gegenbeispiele für die zur jeweiligen Zeit akzeptierte Version des Polyedersatzes gefunden wurden. Wenn man grob überschlägt, wie lange die Beschreibung des menschlichen Gehirns wäre, so kommt man auf eine Größenordnung von etwa $10^{14} = 100\ 000\ 000\ 000\ 000$ oder mehr Buchstaben.[34] Daß eine solch lange Beschreibung des zugrunde liegenden Algorithmus nicht von einem Mathematiker erfaßt werden kann, um daraus den zugehörigen unableitbaren arithmetischen Ausdruck zu bestimmen,

[30] Siehe Penrose [Pen89] Seite 418.

[31] In Penrose [Pen89] heißt es dazu: 'We must *see* - at least in principle - that each step in an argument can be reduced to something simple and obvious.'

[32] Diese 'individuelle' algorithmische Information kann dabei zwar für eine ganze - eventuell auch unendliche - Klasse von Theoremen gelten, jedoch gibt es immer noch unendlich viele solcher Klassen, die jeweils eigene Information benötigen.

[33] Siehe hierzu Lakatos [Lak85].

[34] Bei 10^{11} Neuronen, die bis zu 10000 Verbindungen haben, ergeben sich bereits $\frac{1}{2} \times 10^{15}$ zu beschreibende Synapsen. Wenn jede Synapse durch zwei Neuronen, die sie verbindet, charakterisiert wird, so benötigt man für jede Synapse zwei Zahlen von 11 Ziffern. Selbst wenn diese Berschreibung in hohem Maße redundant ist, z.B. nur der Bruchteil von $\frac{1}{100\ 000\ 000}$ essentiell ist, so verbliebe immer noch eine Textlänge von etwa 110 000 000 Ziffern übrig.

versteht sich von selbst.[35] Dabei wird der zu bestimmende Ausdruck sogar noch unvorstellbar viel länger werden.[36] Daß man bei einem solch langen Ausdruck nicht daran denken kann, daß ein Mensch die Wahrheit des Ausdrucks *einsehen* kann, ist wohl offensichtlich. Auf der anderen Seite zeigt gerade der Gödelsche Beweis, insbesondere Chaitins informationstheoretische Version von Gödels Beweis[37] - daß sich ein Beweis mit wenigen formalen Mitteln nicht führen läßt. Somit ist es also auch nicht möglich, daß beispielsweise Generationen von Mathematikern - zumindest im Prinzip[38] - sukzessive die Wahrheit einer solchen langen Aussage nach allgemein akzeptierten und festgeschriebenen Regeln beweisen könnten !
Damit wäre also dargelegt, daß Argumentationen, die mit dem Gödelschen Unvollständigkeitstheorem für den nicht-algorithmischen Charakter des menschlichen Bewußtseins argumentieren, nicht durchschlagend sind.

[35] Wenn der Mathematiker eine Ziffer in $\frac{1}{10}$ Sekunde liest, so würde er 3 170 979 Jahre dafür benötigen, nur die Beschreibung des Algorithmus zu lesen. Nicht nur, daß er derweil schon längst verstorben wäre, er würde sicher auch fast alles Gelesene sofort wieder vergessen ! Man kann sich wohl schwer vorstellen, daß ein Mathematiker auch nur 'im Prinzip' - was immer das hier heißen mag - noch irgendein Verständnis für die vorliegende Beschreibung aufbringen könnte. Gar nicht zu denken ist an die Zeit die er zur Niederschrift des daraus zu konstruierenden arithmetischen Ausdrucks, der noch unvorstellbar viel länger wäre, benötigen würde !

[36] Siehe Nagel&Newman [NN64] für eine Erläuterung des Gödelschen Beweises.

[37] Vergleiche Abschnitt 3.4.

[38] Was soll hier 'im Prinzip' überhaupt heißen ? Daß ich könnte, wenn ich nicht müde werden würde, keine Magenverstimmung bekäme, immer motiviert wäre, nicht sterben müßte, etc. Vielleicht gehören diese 'Störfaktoren' ja genauso zu meiner Natur, wie meine kognitiven Fähigkeiten. Und selbst wenn all diese Störfaktoren fehlen würden, wären meine kognitiven Kapazitäten nicht trotzdem beschränkt ? Vielleicht ähnlich, wie ein endlicher Automat zwei beliebig große Zahlen nur addieren - nicht aber multiplizieren kann. Für die Multiplikation müßte er eine unbegrenzte Zahl von verschiedenen internen Zuständen einnehmen können, was er per definitionem nicht kann. Kann ich es ?

Kapitel 10

Zusammenfassung und Schlußfolgerungen

Im folgenden sollen die wichtigsten Punkte der vorhergehenden Kapitel noch einmal angesprochen werden und daraus resultierende Schlußfolgerungen - einerseits für eine Methodologie einer künstlichen Intelligenz und andererseits für die Philosophie - gezogen werden.
Der zentrale Begriff der Arbeit war der Begriff der Kolmogoroffkomplexität oder - synonym - der algorithmischen Information, welcher ein *Beschreibungs*komplexitätsmaß ist. Es wird gemessen, wie lang eine Beschreibung gegebener Phänomene *mindestens* sein muß - unabhängig davon, ob man die Phänomene durch Allgemeinaussagen, durch Konstruktionsvorschriften oder durch eine Vielzahl von Regeln, die gegenseitig voneinander abhängen beschreibt. Letzteres kommt dem Konnektionismus entgegen, der sich nicht nur durch technische Neuerungen, sondern auch aufgrund philosophischer Argumente seit einigen Jahren einem zunehmenden Interesse sowohl in der Philosophie, der Kognitionswissenschaft als auch in der Informatik erfreut.
In Abschnitt 8.4 wurde der Konnektionismus aus der Perspektive der Kolmogoroffkomplexität betrachtet. Dabei wurde betont, daß mögliche Intelligenzleistungen konnektionistischer Systeme im wesentlichen auf die Strukturkomplexität, d.h. die Beschreibungskomplexität, des Gesamtsystems zurückgeführt werden müssen. Damit konnte aufgezeigt werden, daß die konnektionistische Idee als Grundlage einer Kognitionswissenschaft, wie es etwa Smolensky [Smo88] fordert, keine merkliche Annäherung an die komplexen Phänomene darstellt, von denen eine Kognitionswissenschaft handeln soll.
Die auf den ersten Blick so überzeugende Nähe zur biologischen Funktionsweise des menschlichen Gehirns einerseits und die dynamischen Systemeigenschaften, die der Wittgensteinschen Regelanwendung und Regelfortschreibung andererseits zu entsprechen scheinen, täuscht leicht über die wirklichen Probleme einer Kognitionswissenschaft - die hohe Komplexität menschlicher Kognitionen im Sinne einer Beschreibung - hinweg. Wie in Hoffmann [Hof91b] aufgezeigt wurde, gehen auch Anstrengungen, die biologische Funktionsweise menschlicher oder tierischer Neuronen möglichst genau zu modellieren, an dem Problem, menschliche *Kognitionen* zu beschreiben oder zu erklären, im wesentlichen vor-

bei.
Der Hauptanteil an der Erklärung des emergenten Verhaltens eines großen konnektionistischen Netzwerkes - wie etwa dem menschlichen Gehirn - geht weniger auf die detaillierte Funktionsweise einzelner Neuronen zurück, sondern auf die spezifische, hochkomplexe Topologie des Netzwerkes.
Experimente mit konnektionistischen Systemen, menschliche kognitive Leistungen - zumindest im kleinen Maßstab - nachzubilden, werden dadurch leicht fehlinterpretiert. Der Erfolg solcher Experimente läßt keine Verallgemeinerung auf die besondere Eignung konnektionistischer Modelle zur Erklärung oder Simulation menschlicher Kognitionen zu.
Dieser Fehlschluß, der gleichermaßen vielen nicht-konnektionistischen, sogenannten *symbolischen* Ansätzen zur Kognitionswissenschaft und der künstlichen Intelligenz zugrunde liegt, wurde in Abschnitt 6 behandelt.
Dort wurde aufgezeigt, daß aufgrund der sehr hohen Beschreibungskomplexität menschlicher Kognitionen, die Gefahr besteht, Experimente zu entwerfen, die nur einen sehr kleinen Teil der insgesamt notwendigen Beschreibungskomplexität erfordern. Dadurch ist es möglich, einfache Erklärungsmodelle für das jeweilige Experiment zu finden. Jedoch ist es typisch, daß diese einfachen Erklärungsmodelle sich nicht auf umfassendere Phänomenbereiche übertragen lassen. In diesem Zusammenhang wurde auf die Bedeutung von universalistischen Vorstellungen menschlicher Kognitionen hingewiesen. Die universalistischen Vorstellungen sind verantwortlich für den spezifischen Aufbau eines Experiments und der Einschätzung der Repräsentativität des Experiments für die Gesamtheit menschlicher Kognitionen.
Damit konnte in der Arbeit mit Hilfe des Begriffs der Kolmogoroffkomplexität aufgezeigt werden, inwiefern für Wissenschaften mit einem hochkomplexen Gegenstandsbereich, wie die Kognitionswissenschaft und die künstliche Intelligenz, eine andere Methodologie in Ansatz gebracht werden muß, als in Wissenschaften wie der Physik, in denen der Gegenstandsbereich von vergleichsweise geringer Komplexität ist.
An dieser Stelle stellt sich allerdings die Frage, wie die Forderung nach einer anderen, neuen Methodologie in Kognitionswissenschaft und KI inhaltlich erfüllt werden kann.
Hierzu scheint die phänomenologische Kritik, aber auch die Analyse des menschlichen Regelfolgens eine wichtigen Beitrag zu leisten. Die generelle Kritik an der Möglichkeit einer künstlichen Intelligenz aus der Perspektive der Phänomenologie Heideggers trifft in ihrer radikalen Form[1] nicht nur die KI selbst, sondern auch eine Kognitionswissenschaft, die auf *operationale* Beschreibungen menschlicher Denkprozesse abzielt.
Die phänomenologische Kritik an der generellen Möglichkeit einer künstlichen Intelligenz wurde in Kapitel 7 zurückgewiesen. Ihre Zurückweisung bietet damit auch eine Grundlage für die Möglichkeit einer symbolisch orientierten Kognitionswissenschaft, wie sie unter anderem von Fodor & Pylyshyn in [FP88] gefordert wird.
Letztlich gründet die Zurückweisung der phänomenologischen Kritik darin, daß die phänomenologische Perspektive auf der Basis eines sehr beschränkten Bewußtseins beruht, das versucht, hochkomplexe kognitive Prozesse zu beobachten und zu erklären. Diese Basis hat bei

[1] Wie sie von Dreyfus in [Dre72] vorgetragen wurde. Diese steht im Gegensatz zu späteren Arbeiten, in denen Dreyfus die Möglichkeit einer konnektionistischer KI einräumt - z.B. in [DD87].

KI-Systemen oder formalen Beschreibungen - wie sie im Falle der Kognitionswissenschaft zur Debatte stehen - kein Pendant.
Daß sich bereits aufgrund dieser beiden genannten Bedingungen - das beschränkte Bewußtsein und die komplexen kognitiven Prozesse - eine Heideggers Analyse entsprechende Erscheinung der menschlichen Kognitionen im Bewußtsein ergibt, wurde in Abschnitt 8.3 aufgezeigt.
Ein Problem, das ebenfalls mit einem beschränkten Bewußtsein zu tun hat, wurde bei der Frage nach Kreativitätsleistungen aufgezeigt. In Abschnitt 8.6 wurde die Anwendung des Begriffs der topischen Kreativität auf Maschinen als Kategorienfehler reklamiert.
Die Nichtanwendbarkeit des (topischen) Kreativitätsbegriffs auf Maschinen bedeutet dabei allerdings nicht, daß Maschinen grundsätzlich nicht zu einer vergleichbaren Leistung imstande sind. Ein vergleichbares Ergebnis ist im Prinzip erzielbar, wenn auch der Weg dorthin ein völlig anderer ist, weil Computern kein Bewußtsein zugeschrieben wird. Dieses jedoch wurde als erforderlich für die Unterscheidung zwischen topischer und kombinatorischer Kreativität herausgestellt. Inwiefern man allerdings eine von Computern scheinbar hervorgebrachte (topische) Kreativitätsleistung nicht der Maschine, sondern dem Programmierer oder Entwickler zuschreiben kann oder muß, ist zunächst noch eine offene Frage. Scheinen doch die Grenzen zwischen kombinatorischer und topischer Kreativität mit zunehmender Systemkomplexität ihre Klarheit zu verlieren.
Als eine andere Konsequenz zunehmender Systemkomplexität kann letztlich auch der Wittgenstein'sche Regelbegriff gesehen werden. Wie in Abschnitt 8.3 ausgeführt, hat eine komplexe Regel bei dem Versuch, sie einfach zu erfassen, genau die von Wittgenstein beobachteten Eigenschaften: Die Regel hat - einerlei wie man sie formuliert - immer nicht näher angebbare Ausnahmen.
Der interessantere Aspekt von Wittgensteins Untersuchungen scheint in dieser Hinsicht aber die Regel*fortschreibung* zu sein.
Hierbei genügt es für eine künstliche Intelligenz sicher nicht, festzustellen, *daß* eine Regelfortschreibung, eine Veränderung, Weitung oder Verengung einer Regel stattfindet. Hingegen ist es für eine künstliche Intelligenz entscheidend, wie eine Regel in jedem konkreten Einzelfall fortgeschrieben wird.
Nun schließt sich der Kreis; ich komme zurück zu den methodologischen Problemen einer KI oder Kognitionswissenschaft.
Sowohl der späte Wittgenstein als auch Heideggers Phänomenologie weisen nicht nur darauf, daß die kognitiven Prozesse von sehr hoher Kolmogoroffkomplexität sind, sondern daß gerade die Regelfortschreibung ebenfalls von erheblicher Komplexität ist.
In diesem Zusammenhang sollen die drei folgenden Stufen von Wissen unterschieden werden:

1. Beziehungen zwischen Objekten oder Begriffen, die sich modelltheoretisch beschreiben lassen. Hierfür läßt sich z.B. die Prädikatenlogik nutzen.

2. Weiteres 'Wissen', welches sich zwar noch auf eine gegenwärtige statische Zustandsbeschreibung einer 'äußeren Welt' bezieht, die aber nicht unter 1. fällt. Dazu

zählt unsicheres Wissen, unvollständiges Wissen, Präferenzen unter konkurrierenden Hypothesen, etc. Für diese Wissensarten wurden und werden in der künstlichen Intelligenz verschiedene Repräsentations- und Schlußformalismen entwickelt.

3. 'Wissen' das beschreibt, wie Regeln, Begriffe, oder Wissen der ersten beiden Stufen fortgeschrieben werden. Dies bezieht sich damit auf die dynamischen Aspekte von Wissen auf der ersten und zweiten Stufe (und eventuell auch auf der dritten Stufe).

Während die erste Stufe sich verhältnismäßig leicht durch Introspektion[2] oder Analyse von Fachtermini explizieren läßt, trifft man bei der zweiten Stufe schon auf mehr Schwierigkeiten.
Bei Wissen der dritten Stufe erscheint es unklar, ob es introspektiv überhaupt erfaßt werden kann. Aber für eine weiterführende künstliche Intelligenz wäre die Explizierung solcher Regeln unabdingbar. Mithin weist die phänomenologische Kritik an der symbolischen künstlichen Intelligenz darauf hin, daß eine KI Methoden entwickeln muß, um entweder menschliches Wissen dieser Art explizieren zu können oder aber Wissen dieser Art anderweitig zu rekonstruieren.
Bereits einzelne Exemplifizierungen dieses Wissens anhand von faktischen Regelfortschreibungen scheinen nur in beschränkten Maße möglich zu sein (z.B. beim induktiven Schließen).
Wenn aber die Annahme zutrifft, daß die dritte Stufe eine erhebliche Kolmogoroffkomplexität aufweist, so würde dies auch bedeuten, daß einzelne Exemplifizierungen nicht viel nutzen, um die zugrundeliegenden konkreten Regeln zur Regelfortschreibung zu gewinnen. Dies würde wohl auch die praktische Undurchführbarkeit von Pylyshyns Vorschlag zur Grundlegung einer Kognitionswissenschaft implizieren. Wenn nämlich das 'Wissen' der zweiten und dritten Stufe von hoher Komplexität ist, so bedeutet das für Pylyshyn zumindest, daß seine *funktionale Architektur* von hoher Komplexität ist - denn diese Regeln in ihrer reinen Form, das heißt abgesehen von irgendwelchen konkret involvierten Inhalten, gehören ganz sicher zu seiner funktionalen Architektur. Festzustellen welche Form dieser Regeln die reine, also die kognitiv unbeeinflußbare Form ist, wird auf erhebliche Schwierigkeiten stoßen, wenn die Inhalte des 'Wissens' der dritten Stufe nicht introspektiv zugänglich sind. Noch viel schwieriger ist es dann, festzustellen ob der Inhalt sich durch kognitive Faktoren verändert !

Eine wichtige erkenntnistheoretische Frage, die hier auftaucht, ist die Folgende:
Inwieweit können (algorithmische) Regeln der Regelfortschreibung auf andere Weise rekonstruiert werden, als durch Introspektion ?
Die künstliche Intelligenz interessiert sich ohnehin eher dafür, wie die Regeln aussehen *sollten*, als dafür, wie sie bei einem einzelnen Indiviuum tatsächlich aussehen. Daher sind für die KI die Möglichkeiten einer Rekonstruktion von besonderem Interesse.

[2] Mit *Introspektion* ist hier eine Selbstbeobachtung der im Bewußtsein stattfindenden kognitiven Prozesse gemeint, mit der Intention, diese Prozesse erklären zu wollen.

Weiterhin bietet der Begriff der Kolmogoroffkomplexität unter anderem für die folgenden philosophischen Problembereiche neue Perspektiven.
Die Feststellung eines philosophischen Regelcharakters, z.B. bei der menschlichen Begriffs- und Sprachverwendung, erscheint vor dem Hintergrund des Begriffs der Kolmogoroffkomplexität als unnötig restriktiv. Diese nur qualitative Einordnung liegt auch Dreyfus' Kritik an der KI zugrunde. Durch die Einführung des quantitativen Aspektes konnte in Kapitel 9 die Frage nach den generellen Grenzen der KI einer solchen Präzisierung zugeführt werden, daß sie zumindest theoretisch beantwortet werden könnte.

Insbesondere im Hinblick auf die offensichtliche Komplexitätsbeschränkung der bewußt wahrnehmbaren Denk- und Reflexionsprozesse erscheinen Betrachtungen der folgenden Art interessant:

- Durch welche Reflexionsstrategien kann eine (oder eine wie große) Komplexitätsdifferenz zwischen bewußten und unbewußten Denkprozessen überwunden werden, so daß alle Denkprozesse expliziert werden können ?
- Wie kann der Kreativitätsbegriff mittels des Begriffs der Kolmogoroffkomplexität geschärft werden ?
- Welche Rolle spielt die Komplexität bei kreativen Leistungen oder bei Metaphern ? Hat eine größere Komplexitätsdifferenz zwischen dem bisherigen Sprachgebrauch und einer neuen Metapher eine andere Qualität zur Folge ? Inwieweit kann eine solche Metapher überhaupt noch verstanden werden ?
- Welche Rolle spielt die Komplexität bei hermeneutischen Prozessen ? Gibt es komplexitätsbedingte Grenzen des Verstehens ? Wo liegen diese Grenzen und welche Eigenschaften des Bewußtseins, welche Reflexionskapazitäten sind dafür verantwortlich ?
- etc. etc.

Die komplexitätstheoretischen Zweifel an der These von Maturana und Varela aus Abschnitt 8.5 über das Zustandekommen intelligenter kognitiver Systeme scheint besonders interessant für weitere Forschung in der theoretischen Informatik und deren Konsequenzen für die Philosophie zu sein. Sollten sich die Hinweise auf unüberwindliche Komplexitätsschranken erhärten, so könnte dies von großer Bedeutung für die gegenwärtige Diskussion um die Verteilung von Intelligenz, Sprachverstehen und anderes auf die phylogenetische bzw. ontogenetische Entwicklung werden. Denn während der ontogenetischen Entwicklung ist das Individuum zum großen Teil auf seine eigene Beurteilung seiner Klassifikationsleistung angewiesen - es bekommt nicht ständig eine Rückmeldung, ob etwas richtig verstanden wurde oder ob eine Handlung erfolgreich war. Insofern scheint ein erheblicher Teil der ontogenetischen Entwicklung auf einer Selbstorganisationsfähigkeit zu beruhen.

Sollte sich beispielsweise herausstellen, daß die phylogenetische Entwicklung den allergrößten Teil des Sprachverstehen bestimmt - also die während des kindlichen und erwachsenen Lebens erlernten Sprachverwendungregeln und das Sprachverstehen insgesamt nur einen fast vernachlässigbaren Einfluß haben, so könnte dies die Quine'sche These der generellen Nichtübersetzbarkeit erschüttern.

Würde sich hingegen herausstellen, daß unter ganz bestimmten Voraussetzungen die Entwicklung zu komplexen kognitiven Systemen durch Selbstorganisation möglich ist, so könnte dies einen starken Hinweis auf die tatsächlich (dann notwendigerweise) vorherrschenden Entwicklungsbedingungen geben.
Erkenntnisse solcher Art könnten Einfluß auf einen weiten Bereich philosophischer Fragestellungen haben.

Literaturverzeichnis

[Adl79] L. Adleman. Time, space and randomness. Technical report, Massachusetts Institute of Technology, March 1979. MIT/LCS/79/TM-131.

[AEM87] W. Arnold, H. J. Eysenck, and R. Meili. *Lexikon der Psychologie*, volume 2. Herder, 1987.

[Amb91] S. Ambroskiewicz. From chaos to knowledge. In *Proceedings of the World Conference on the Fundamentals of Artificial Intelligence*, pages 37–43, 1991.

[Arb89] M. A. Arbib. *The Metaphorical Brain II.* Wiley, 1989.

[Ari] Aristoteles. *Metaphysik.*

[Ash47] W. R. Ashby. Principles of the self-organizing dynamic system. *Journal of General Psychology*, 37, 1947.

[Bab70] C. Babbage. Of the analytical engine (1864). In Z. W. Pylyshyn, editor, *Perspectives on the Computer Revolution*, pages 16–28. Prentice-Hall, Englewood Cliffs, NJ, 1970.

[BC91] A. Bauval and L. Cholvy. Automated reasoning in case of inconsistency. In *Proceedings of the World Conference on the Fundamentals of Artificial Intelligence*, pages 81–92, 1991.

[Ben67] P. Benacerraf. God, the devil, and Gödel. *Philosophy*, 36:112–117, 1967.

[Ber10] G. Berkeley. *Treatise concerning the Principles of Human Knowledge.* 1710.

[BH64] Y. Bar-Hillel. The present status of automatic translation of language. In F. L. Alt, editor, *Advances in Computers*, volume 1. Academic Press, New York, 1964.

[BL85] R. J. Brachman and H. J. Levesque, editors. *Readings in knowledge representation.* Morgan Kaufmann, Los Altos, CA, 1985.

[Bla62] M. Black. *Models and Metaphors.* Cornell University Press, Ithaca, N.Y., 1962.

[Bla84] S. Blackburn. The Individual Strikes Back. *Synthese*, 84:281–301, 1984.

[Blo78] N. Block. Troubles with functionalism. In C. W. Savage, editor, *Minnesota studies in philosophy of science*, volume 9. Minneapolis, MN, 1978.

[Blo90] N. Block. The computer model of the mind. In D. N. Osherson and E. E. Smith, editors, *Thinking: An invitation to cognitive science*, volume 3, pages 247–289. MIT Press, 1990.

[BP90] J. Briggs and F. D. Peat. *Die Entdeckung des Chaos.* Hanser-Verlag, 1990.

[Bra90] R. J. Brachman. The future of knowledge representation. In *Proceedings of the* 8^{th} *Conference of the American Association of Artificial Intelligence*, pages 1082–1092, 1990.

[Bri82] L. Briskman. Creative product and creative process in science and art. In D. Dutton and M. Krausz, editors, *The concept of creativity in Science and Art*, pages 129–155. Martin Nijhoff Publishers, 1982.

[Bro91] R. Brooks. Intelligence without reason. In *Proceedings of the* 12^{th} *International Joint Conference on Artificial Intelligence*, pages 569–595, 1991.

[BS84] B. G. Buchanan and E. H. Shortliffe. *Rule-Based Expert Systems.* Addison-Wesley, 1984.

[Bur80] H. Burkhardt. *Logik und Semiotik in der Philosophie von Leibniz.* München, 1980.

[Car28] R. Carnap. *Der logische Aufbau der Welt.* 1928.

[Car50] R. Carnap. *Logical Foundations of Probability.* 1950.

[CC50] W. G. Cochran and G. M. Cox. *Experimental Design.* New York, 1950.

[Cha66] G. J. Chaitin. On the length of programs for computing finite binary sequences. *Journal of the Association of Computing Machinery*, 13:547–569, 1966.

[Cha74] G. J. Chaitin. Information-theoretic Limitations of Formal Systems. *Journal of the ACM*, 21, 1974.

[Cha87] G. J. Chaitin. *Algorithmic information theory.* Cambridge University Press, 1987.

[Che78] V. Cherniavsky. On algorithmic natural language analysis and understanding. *Information Systems*, 10:5–10, 1978.

[Cho77] N. Chomsky. *Reflexionen über die Sprache.* Suhrkamp, Frankfurt a.M., 1977. (Engl. Orig.: *Reflections on Language.* Pantheon, New York, 1975.).

[Chu36a] A. Church. A note on the Entscheidungsproblem. *Journal of Symbolic Logic*, 1:40–41,101–102, 1936.

[Chu36b] A. Church. An unsolvable problem of elementary number theory. *American Journal of Mathematics*, 58, 1936.

[Coo71] S. A. Cook. The complexity of theorem-proving procedures. In *Proceedings of the* 3^{rd} *Annual ACM Symposium on Theory of Computing*, pages 151–158, 1971.

[Cra92] H. J. Cram. Fodor's causal theory of representation. *Philosophical Quarterly*, 42(166):56–70, January 1992.

[Cum89] R. Cummins. *Meaning and mental representation.* MIT Press, Cambridge, 1989.

[Dav82] M. Davis. Why Gödel didn't have Church's thesis. *Information and Control*, 54, 1982.

[DD87] H. L. Dreyfus and S. E. Dreyfus. *Künstliche Intelligenz. Von den Grenzen der Denkmaschinen und dem Wert der Intuition.* Rowohlt Verlag, 1987. (Engl. Orig.: *Mind over Machine.* Free Press, New York, 1986.

[DD88] H. L. Dreyfus and S. E. Dreyfus. Making a mind versus modelling a brain: Artificial intelligence back at a branchpoint. In S. R. Graubard, editor, *The Artificial Intelligence Debate*, pages 15–43. MIT Press, 1988.

[Den71] D. C. Dennett. Intentional systems. *Journal of Philosophy*, 68:87–106, 1971.

[Den87] D. C. Dennett. *The Intentional Stance.* MIT Press, 1987.

[Die86] T. G. Dietterich. Learning at the knowledge level. *Machine Learning*, 1:287–316, 1986.

[Die91] J. Diederich. Trends im Konnektionismus. *Künstliche Intelligenz*, 2:6–11, 1991.

[DM64] R. A. Dentler and B. Mackler. Originality. *Behavior Science*, 9:1–10, 1964.

[Dre65] H. L. Dreyfus. Alchemy and artificial intelligence. Technical Report P3244, RAND Corporation, Santa Monica, CA, December 1965.

[Dre72] H. L. Dreyfus. *What Computers Can't do - The Limits of Artificial Intelligence.* Harper & Row, 1972.

[Dre86] F. Dretske. Misrepresentation. In R. J. Bogdan, editor, *Belief.* Oxford, 1986.

[Eig71] M. Eigen. Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58, 1971.

[FFGL90] J. A. Feldman, M. A. Fanty, N. H. Goddard, and K. J. Lynne. Computing with structured connectionist networks. In S. F. Zornetzer, J. L. Davis, and C. Lau, editors, *An Introduction to Neural and Electronic Networks*, pages 433–454. Academic Press, New York, 1990.

[FH92] R. Freivalds and A. G. Hoffmann. An inductive inference approach to classification. In *Proceedings of the* 3^{rd} *Workshop on Analogical and Inductive Inference*, Schloß Dagstuhl, Germany, October 1992. Springer-Verlag. (Erscheint demnächst.).

[Fis60] R. A. Fisher. *The Design of Experiments.* Edinburgh, 1960.

[FK77] R. Freivalds and E. Kinber. Limiting identification of the minimal Gödel numbers. *Theory of algorithms and programs*, 3:3–34, 1977. (Russian.).

[FM90] J. A. Fodor and B. P. McLaughlin. Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work. *Cognition*, 35:183–204, 1990.

[Fod75] J. A. Fodor. *The language of thought.* Harvard University Press, Cambridge, MA, 1975.

[Fod78] J. A. Fodor. Tom Swift and his procedural grandmother. *Cognition*, 6:229–247, 1978.

[Fod80] J. A. Fodor. Methodological solipsism as a research strategy in psychology. *Behavioral and Brain Sciences*, 3:63–73, 1980.

[Fod87a] J. A. Fodor, editor. *Psychosemantics.* MIT Press/Bradford Books, 1987.

[Fod87b] J. A. Fodor. Why there still has to be a language of thought. In J. A. Fodor, editor, *Psychosemantics.* MIT Press/Bradford Books, 1987.

[FP88] J. A. Fodor and Z. W. Pylyshyn. Connectionism and cognitive architecture: a critical analysis. *Cognition*, 28:3–71, 1988.

[FR74] M. J. Fischer and M. O. Rabin. Super-exponential complexity of Presburger arithmetic. In R. M. Karp, editor, *Complexity of Computation*, pages 27–41. American Mathematical Society, 1974.

[Fre79] G. Frege. *Begriffsschrift, eine der arithmetischen nachgebildeten Formelsprache des reinen Denkens.* Halle, 1879. (Neudr. in: K. Berka & L. Kreiser (eds.) Logik-Texte, Berlin (Ost), 1971).

[Fre90] R. M. French. Subcognition and the limits of the Turing test. *Mind*, 99:53–65, 1990.

[Fri91] M. Frixione. On the relations between philosophical theories of meaning and Artificial Intelligence. In *Proceedings of the World Conference on the Fundamentals of Artificial Intelligence*, pages 187–198, 1991.

[FSG89] M. Frixione, G. Spinelli, and S. Gaglio. Symbols and subsymbols for representing knowledge: a catalogue raisonné. In *Proceedings of the* 11^{th} *International Joint Conference of Artificial Intelligence*, pages 3–7, 1989.

[Göd31] K. Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 39, 1931.

[Goo46] N. Goodman. A query of confirmation. *Journal of Philosophy*, 43:383–385, 1946.

[Goo55] N. Goodman. *Fact, Fiction and Forecast.* Cambridge, MA, 1955. (dt. Tatsache, Fiktion, Voraussage; Frankfurt, 1975).

[Goo69] N. Goodman. *Languages of Art.* Oxford University Press, London, 1969.

[GQ47] N. Goodman and W. v. O. Quine. Steps toward a constructive nominalism. *Journal of Symbolic Logic*, 12:105–122, 1947.

[HACN90] F. Hsu, T. Anantharaman, M. S. Campbell, and A. Nowatzyk. A grandmaster chess machine. *Scientific American*, 4(263):44–50, 1990.

[Han87] S. Hanard. Category induction and representation. In S. Hanard, editor, *Categorial Perception*, pages 535–565. Cambridge University Press, 1987.

[Har87] G. Harman. (Nonsolipsistic) conceptual role semantics. In E. Le Pore, editor, *New directions in semantics.* Academic Press, New York, 1987.

[Har91] S. Harnad. Other bodies, other minds: A machine incarnation of an old philosophical problem. *Minds and Machines*, 1(1):43–54, 1991.

[Hau78] J. Haugeland. The nature and plausibility of cognitivism. *Behavioral and Brain Sciences*, 2:215–260, 1978.

[HB68] D. Hilbert and P. Bernays. *Grundlagen der Mathematik I.* Springer-Verlag, 1968. 2. Auflage.

[HD72] B. Hoffmann and H. Dukas. *Albert Einstein: Creator and Rebel.* New York, 1972.

[Heb49] D. O. Hebb. *The organization of behavior.* Wiley & Sons, 1949.

[Hei27] M. Heidegger. *Sein und Zeit.* 1927.

[Hei70] J. v. Heijenoort. *Frege and Gödel. Two Fundamental Texts in Mathematical Logic.* Harvard University Press, Cambridge, Massachussetts, 1970.

[Hin89] G. E. Hinton. Connectionist learning procedures. *Artificial Intelligence*, 40:185–234, 1989.

[Hob51] T. Hobbes. *Leviathan*. 1651.

[Hof85] D. R. Hofstadter. Waking up from the boolean dream, or, subcognition as computation. In D. R. Hofstadter, editor, *Metamagical themas*, pages 631–665. Basic Books, 1985.

[Hof90a] A. G. Hoffmann. General limitations on machine learning. In *Proceedings of the* 9^{th} *European Conference on Artificial Intelligence*, pages 345–347, Stockholm, Sweden, August 1990.

[Hof90b] A. G. Hoffmann. On computational limitations of neural network architectures. In *Proceedings of the* 2^{nd} *IEEE Symposium on Parallel and Distributed Processing*, pages 818–825, Dallas, Texas, USA, December 1990. IEEE.

[Hof91a] A. G. Hoffmann. Asymptotic performance of learning algorithms. In *Proceedings of the first World Conference on the Fundamentals of Artificial Intelligence*, pages 247–256, 1991.

[Hof91b] A. G. Hoffmann. Connectionist functionality and the emergent network behavior. *Neurocomputing - An International Journal*, 1991. to appear.

[Hof91c] A. G. Hoffmann. On the principles of intelligence. In *Proceedings of the first World Conference on the Fundamentals of Artificial Intelligence*, pages 257–266, 1991.

[Hof92] A. G. Hoffmann. Phenomenology, representations and complexity. In *Proceedings of the* 10^{th} *European Conference on Artificial Intelligence*, Vienna, Austria, August 1992. Wiley & Sons.

[Hol86] J. H. Holland. Escaping brittleness: The possibilities of general-purpose learning algorithms applied to parallel rule-based systems. In R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, editors, *Machine Learning: An Artificial Intelligence Approach*, volume 2, pages 593–623. Morgan Kaufmann Publishers, 1986.

[Hub78] C. Hubig. *Dialektik und Wissenschaftslogik. Eine sprachphilosophisch-handlungstheoretische Analyse*. de Gruyter, Berlin, New York, 1978.

[Huc89] P. Hucklenbroich. *Künstliche Intelligenz und medizinisches Wissen. Wissenschaftstheoretische Grundfragen von Expertensystemen und wissensbasierter Programmierung*. Universität Münster, Medizinische Fakultät, Münster, 1989. Habilitationsschrift.

[Hum48] D. Hume. *A Treatise of Human Nature*. 1748.

[Hus13] E. Husserl. *Ideen zu einer reinen phänomenologischen Philosophie.* 1913. Bd. I & II.

[Hus85] E. Husserl. *Die phänomenologische Methode. Ausgewählte Texte.* Reclam, Stuttgart, 1985.

[Ind90] B. Indurkhya. Some remarks on the rationality of induction. *Synthese*, 85:95–114, 1990.

[Jac87] D. Jacquette. Metamathematical criteria for minds and machines. *Erkenntnis*, 27:1–16, 1987.

[Jel90] E. Jelden. Menschliche und elektronische Wissensverarbeitung in der Heuristik. Voraussetzungen, Grenzen und Möglichkeiten aus der philosophischen Sicht. Technical report, TU Berlin, Institut für Philosophie, Wissenschaftstheorie und Wissenschafts- und Technikgeschichte, Berlin, Mai 1990.

[Joh87] M. Johnson. *The Body in the Mind: The Bodily Basis of Meaning, Imagination and Reason.* University of Chicago Press, 1987.

[Kan87] I. Kant. *Kritik der reinen Vernunft.* Suhrkamp, Frankfurt, 1787. stw 55.

[Kir92] W. W. Kirchherr. Kolmogorov complexity and random graphs. *Information Processing Letters*, 41:125–130, March 1992.

[Kö68] S. Körner. *Philosophie der Mathematik. Eine Einführung.* Nymphenburger, München, 1968.

[Koe82] A. Koestler. Three domains of creativity. In D. Dutton and M. Krausz, editors, *The concept of creativity in Science and Art*, pages 1–17. Martin Nijhoff Publishers, 1982.

[Kol65] A. N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems in Information Transmission*, 1(1):1–7, 1965.

[Kri59] S. A. Kripke. A completeness theorem in modal logic. *Journal of Symbolic Logic*, 24:1–14, 1959.

[Kri82] S. A. Kripke. *Wittgenstein on Rules and Private Language.* Blackwells, Oxford, 1982.

[Lak85] I. Lakatos. *Beweise und Widerlegungen. Die Logik mathematischer Entdeckungen.* Braunschweig, 1985.

[Lak87] G. Lakoff. *Women, Fire and Dangerous Things. What Categories Reveal about the Mind.* University Press Chicago, 1987.

[LD89] T. E. Lange and M. G. Dyer. High-level inferencing in a connectionist network. *Connection Science*, 1(2):187–217, 1989.

[Lev88a] H. Levesque. Logic and complexity of reasoning. *Journal of Philosophical Logic*, 17:355–389, 1988.

[Lev88b] D. Levy. *Computer Chess Compendium*. Batsford, London, 1988.

[LHM+91] R. Levinson, F.-H. Hsu, T. A. Marsland, J. Schaeffer, and D. E. Wilkins. Panel: The Role of Chess in Artificial Intelligence Research. In *Proceedings of the* 12^{th} *International Joint Conference on Artificial Intelligence*, pages 547–552, 1991.

[Loc90] J. Locke. *An Essay Concerning Human Understanding*. 1690.

[Luc61] J. R. Lucas. Minds, machines, and Gödel. *Philosophy*, 36:112–117, 1961. (rpt. in: *Minds and Machines*, ed. Alan R. Anderson, Englewood Cliffs, N.J., Prentice-Hall, 1964.).

[Luc65] A. S. Luchins. Mechanisierung beim Problemlösen. In C. F. Graumann, editor, *Denken*, pages 171–190. Kiepenheuer & Witsch, Köln, 1965.

[Luc70] J. R. Lucas. *The freedom of the will*. Clarendon Press, Oxford, 1970.

[LV88] M. Li and P. M. B. Vitanyi. Two decades of applied Kolmogorov complexity. In *Proceedings of the* 3^{rd} *Annual Conference on Structure in Complexity Theory*, pages 80–101, 1988.

[LV89] M. Li and P. M. B. Vitanyi. Inductive reasoning and Kolmogorov complexity. In *Proceedings of the* 6^{rd} *Annual Conference on Structure in Complexity Theory*, pages 165–185, 1989.

[LV91] M. Li and P. M. B. Vitanyi. Combinatorics and Kolmogorov complexity. In *Proceedings of the* 6^{rd} *Annual Conference on Structure in Complexity Theory*, pages 154–163, 1991.

[Mah84] B. Mahr. Die Herrschaft der Gebrauchsanweisung. pages 89–107. Kursbuch Verlag GmbH, Berlin, März 1984.

[Mat70] H. R. Maturana. Neurophysiology of cognition. In P. Garvin, editor, *Cognition: A multiple view*. New York, 1970.

[MB88] S. Muggleton and W. Buntine. Machine invention of first-order predicates by inverting resolution. In *Proceedings of the* 5^{th} *Machine Learning Conference*, pages 339–352. Kaufmann, 1988.

[McC79] P. McCorduck. *Machines Who Think*. W. H. Freeman, San Francisco, CA, 1979.

[McC88] J. McCarthy. Mathematical logic in artificial intelligence. In S. R. Graubard, editor, *The Artificial Intelligence Debate*, pages 297–311. MIT Press, 1988.

[Mea89] C. Mead. *Analog VLSI and neural Systems.* Addison & Wesley, 1989.

[MF90] S. Muggleton and C. Feng. Efficient induction of logic programs. In *Proceedings of the International Conference on Algorithmic Learning Theory*, 1990.

[Mic90] R. S. e. a. Michalski. *Machine Learning: An Artificial Intelligence Approach, I, II & III.* Morgan Kaufmann Publishers, 1983, 86, 90.

[Mil43] J. S. Mill. *A System of Logic, Rationalism and Induction. Being a Connected View of the Principles of Evidence, and the Methods of Scientific Investigation, I-II.* London, 1843.

[Min68] M. Minsky. *Semantic Information Processing.* MIT Press, 1968.

[Min71] M. Minsky. *Berechnung: Endliche und unendliche Maschinen.* Berlin Union GmbH, Stuttgart, 1971. Engl. Original: *Computation: Finite and Infinite Machines.* Prentice-Hall, 1967.

[Min72] M. Minsky. Form and content in computer science. *Journal of the ACM*, January 1972.

[Min82] M. Minsky. Why People Think Computers Can't. *AI Magazine*, Fall 1982.

[Min86] M. Minsky. *The Society of Mind.* Simon and Schuster, 1986.

[Moo87] J. Moor. Turing test. In S. C. Shapiro, editor, *The Wiley Encyclopedia of Artificial Intelligence.* Wiley, 1987.

[MP43] W. S. McCulloch and W. Pitts. A logical calculus for the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–143, 1943.

[MP69] M. Minsky and S. Papert. *Perceptrons.* MIT Press, Cambridge, MA, 1969.

[MS88] H. Mandl and H. Spada. *Wissenspsychologie.* Psychologie Verlags Union, München, 1988.

[MS90] A. T. Marsland and J. Schaeffer. *Computers, Chess, and Cognition.* Springer-Verlag, Berlin, 1990.

[Mün90] D. Münch. The early works of Husserl and artificial intelligence. *Journal of the British Society of Phenomenology*, 2(21):107–120, 1990.

[MV80] H. R. Maturana and F. J. Varela. *Autopoiesis and Cognition: The Realization of the Living.* Boston, 1980.

[MV82] H. R. Maturana and F. J. Varela. Autopoietische Systeme. Eine Bestimmung der lebendigen Organisation. In H. R. Maturana, editor, *Erkennen: Die Organisation von Wirklichkeit. Ausgewählte Arbeiten zur biologischen Epistemologie*. Braunschweig, 1982. (Engl. Orig.: *Autopoietic Systems. A Characterization of the Living Organization*, 1975.).

[MV87] H. R. Maturana and F. J. Varela. *The tree of knowledge. The biological Roots of Human Understanding*. New Science Library, 1987.

[New80] A. Newell. Physical symbol systems. *Cognitive Science*, 4:135–183, 1980.

[New82] A. Newell. The knowledge level. *Artificial Intelligence*, 18:87–127, 1982.

[Nie86] F. Nietzsche. *Jenseits von Gut und Böse*. Insel Verlag, Frankfurt a.M., 1886.

[NN64] E. Nagel and J. R. Newman. *Der Gödelsche Beweis*. Oldenburg, Wien, 1964.

[NP90] G. Nicolis and I. Prigogine. *Self-Organization in Non-Equilibrium Systems: From Dissipative Structures to Order Through Fluctuations*. New York, 1990.

[NS63] A. Newell and H. Simon. GPS, a program that simulates human thought. In E. Feigenbaum and J. Feldman, editors, *Computers and Thought*, pages 279–293. McGraw Hill, New York, 1963.

[NS76] A. Newell and H. A. Simon. Computer science as empirical inquiry: Symbols and search. *Communications of the Association for Computing Machinery*, 19:113–126, 1976.

[Nun91] J. Nunn. An den Grenzen des menschlichen Geistes. *Computerschach und Spiele*, 1& 2:41–44 & 38–41, 1991.

[Pap88] S. Papert. One AI or many. In S. R. Graubard, editor, *The Artificial Intelligence Debate*. MIT Press, 1988.

[Pas91] R. Paslack. *Urgeschichte der Selbstorganisation: zur Archäologie eines wissenschaftlichen Paradigmas*. Vieweg, 1991.

[Pau79] W. Paul. Kolmogorov's complexity and lower bounds. In *Proceedings of the* 2^{nd} *International Conference on Fundamentals of Computation Theory*, 1979.

[Pea91] D. Pears. Wittgenstein's Account on Rule-Following. *Synthese*, 87:373–383, 1991.

[Pen89] R. Penrose. *The Emperor's new Mind*. Oxford University Press, 1989.

[Pet90] P. Pettit. The Reality of Rule-following. *Mind*, 99:1–21, January 1990.

[Pla] Platon. *Parmenides*.

[Plu74] R. Plutchik. *Foundations of experimental research.* New York, 1974.

[Pop70] K. R. Popper. Normal science and its dangers. In I. Lakatos and A. Musgrave, editors, *Criticism and the growth of knowledge.* Cambridge University Press, 1970.

[Pos43] E. L. Post. Formal reduction of the general combinatorial decision problem. *American Journal of Mathematics*, 65, 1943.

[Pra55] C. Prantl. *Geschichte der Logik im Abendlande.* Leipzig, 1855.

[Pri53] H. H. Price. *Thinking and Experience.* London, 1953.

[Put75] H. Putnam. *Mind, Language and Reality.* Cambridge University Press, Cambridge, 1975.

[Put86] H. Putnam. Meaning holism. In L. Hahn and P. Schilpp, editors, *The Philosophy of W. v. O. Quine.* Open Court Publishers, La Salle, IL, USA, 1986.

[Put88] H. Putnam. Much ado about not very much. In S. R. Graubard, editor, *The Artificial Intelligence Debate*, pages 269–282. MIT Press, 1988.

[Put91] H. Putnam. *Repräsentation und Realität.* Suhrkamp, 1991. (Engl. Original: Representations and Reality. MIT Press, 1988.).

[Pyl84] Z. W. Pylyshyn. *Computation and cognition: Toward a foundation for cognitive science.* MIT Press/Bradford Books, 1984.

[Qui80] W. v. O. Quine. *Wort und Gegenstand.* Reclam, Stuttgart, 1980. Engl. Original: *Word and Object*, 1960.

[Qui89] W. v. O. Quine. *Die Wurzeln der Referenz.* Suhrkamp, Frankfurt, 1989. Engl. Original: *The roots of reference*, 1974.

[Res80] N. Rescher. *Induction. An Essay on Justification of Inductive Reasoning.* Oxford, 1980.

[RMt86] D. E. Rumelhart, J. L. McClelland, and the PDP Research Group. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition,I & II.* MIT Press, Cambridge, MA, 1986.

[Rob65] J. A. Robinson. A machine-oriented logic based on the resolution principle. *Journal of the ACM*, 12, 1965.

[Rog59] C. R. Rogers. Toward a theory of creativity. In H. H. Anderson, editor, *Creativity and its cultivation.* New York, 1959.

[Rog67] H. Rogers Jr. *Theory of Recursive Functions and Effective Computability.* McGraw-Hill, 1967.

[Ros59] F. Rosenblatt. Two theorems of statistical separability in the perceptron. In *Proceedings of the Symposium on the Mechanization of thought*, pages 421–456, London, 1959. Her Majesty's Stationary Office.

[Roy98] J. Royce. The psychology of invention. *Psychological Review*, 5:113–144, 1898.

[Rus12] B. Russell. *Problems of Philosophy.* London, 1912.

[Ryl49] G. Ryle. *The concept of mind.* Oxford, 1949.

[Sam59] A. L. Samuel. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3:210–229, 1959.

[Sam67] A. L. Samuel. Some studies in machine learning using the game of checkers II. *IBM Journal of Research and Development*, 11:601– 617, 1967.

[SAW91] G. Schreiber, H. Akkermans, and B. Wielinga. On problems with the knowledge level perspective. In *AI and Simulation of Behavior (AISB-91).* Springer-Verlag, 1991. 208–221.

[SC86] A. Sloman and J. Cohen. What sorts of machines can understand the symbols they use ? *Proceedings of the Aristotelian Society*, 60:61–95, 1986.

[Sch13] M. Scheler. Der Formalismus der Ethik und die materiale Wertethik. In *Jahrbuch für Philosophie und phänomenologische Forschung*, pages 405–565 (Teil I) & 21–478 (Teil II). Halle a. d. S., 1913.

[SD90] J. W. Shavlik and T. G. Dietterich, editors. *Readings in Machine Learning.* Morgan Kaufmann Publishers, 1990.

[Sea80] J. R. Searle. Minds, brains and programs. *Behavioral and Brain Sciences*, 3:417–424, 1980.

[Sea82] J. R. Searle. The chinese room revisited. *Behavioral and Brain Sciences*, 5:345–348, 1982.

[Sea84] J. R. Searle. *Minds, Machines, Brains and Science.* 1984. dt.: Geist, Hirn und Wissenschaft; Suhrkamp, Frankfurt 1986.

[Sea87] J. R. Searle. *Intentionalität. Eine Abhandlung zur Philosophie des Geistes.* Suhrkamp, 1987. (Engl. Original: Intentionality. An essay in the philosophy of mind. Cambridge University Press, 1983.

[Sle87] P. Slezak. Gödel's theorem and the mind. *Brit. J. Phil. Sci.*, 33:41–52, 1987.

[Smi85] B. C. Smith. Prologue to 'Reflection and semantics in a procedural language'. In R. J. Brachman and H. J. Levesque, editors, *Readings in knowledge representation.* Morgan Kaufmann, Los Altos, CA, 1985.

[Smo87] P. Smolensky. Connectionist AI, symbolic AI, and the brain. *AI Review*, 1:95–109, 1987.

[Smo88] P. Smolensky. On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 11:1–74, 1988.

[Smo90] P. Smolensky. In defence of PTC. *Behavioral and Brain Sciences*, 13(2):407–412, 1990.

[SN58] H. A. Simon and A. Newell. Heuristic Problem Solver: The Next Advance in Operations Research. *Operations Research*, 6, January/February 1958.

[Sol64] R. J. Solomonoff. Complexity-based induction systems: comparisons and convergence theorems. *Information and Control*, 7:1–22 and 224–254, 1964.

[Sti83] S. Stich. *From folk psychology to cognitive science: The case against belief.* MIT Press, Cambridge, 1983.

[Sto86] D. Stove. *The Rationality of Induction.* Clarendon Press, Oxford, 1986.

[SZ90] R. Serra and G. Zanarini. *Complex Systems and Cognitive Processes.* Springer-Verlag, Berlin, 1990.

[Tar36] A. Tarski. Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica*, 1:261–405, 1936.

[Tay64] C. W. Taylor. *Creativity: Progress and Potential.* McGraw-Hill, 1964.

[Tur37] A. M. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 2(42):230–265 and (43) 544–546, 1937.

[Tur50] A. M. Turing. Computing machinery and intelligence. *Mind*, 59:433–460, 1950.

[Tur53] A. M. Turing. Digital computers applied to games. In B. V. Bowden, editor, *Faster than Thought*, pages 286–310. Pitman, 1953.

[Val84a] L. Valiant. Deductive learning. *Philosophical Transactions of the Royal Society of London*, 312, 1984.

[Val84b] L. Valiant. A theory of the learnable. *Communications of the ACM*, 27:1134–1142, 1984.

[Var90] F. J. Varela. *Kognitionswissenschaft – Kognitionstechnik.* Suhrkamp, Frankfurt, 1990. stw 882.

[Vin92] R. E. Vinkhuysen. On the non-existence of the knowledge level. In *Proceedings of the* 10^{th} *European Conference on Artificial Intelligence.* Wiley & Sons, 1992. 620–622.

[von60] H. von Förster. On self-organizing systems and their environments. In M. C. Yovits and S. Cameron, editors, *Self-Organizing Systems; Proceedings of an Interdisciplinary Conference*. Oxford, 1960.

[von80] F. von Kutschera. *Sprachphilosophie*. Uni-Taschenbücher 80, 1980.

[vZ62] H. von Förster and G. W. Zopf. *Principles of Self-Organization*. Oxford, 1962.

[Wan74] H. Wang. *From Mathematics to Philosophy*. Routledge & Kegan, London, 1974.

[Web68] J. Webb. Metamathematics and the philosophy of mind. *Philosophy of Science*, 35:156–178, 1968.

[Web80] J. Webb. *Mechanism, Mentalism and Metamathematics: An Essay on Finitism*. D. Reidel Publ., Dordrecht-Boston, 1980.

[WF86] T. Winograd and F. Flores. *Understanding Computers and Cognition: A new Foundation for Design*. Norwood Publisher, 1986.

[Wie84] O. Wiener. Turings Test. Vom dialektischen zum binären Denken. pages 12–37. Kursbuch Verlag GmbH, Berlin, März 1984.

[Win70] P. H. Winston. *Learning structural descriptions from examples*. PhD thesis, MIT, Cambridge, MA, 1970.

[Win72] T. Winograd. *Understanding Natural Language*. Academic Press, New York, 1972.

[Win92] P. H. Winston. *Artificial Intelligence*. Addison & Wesley, 1992. 3^{rd} edition.

[Wit21] L. Wittgenstein. Tractatus logico-philosophicus. In W. Ostwald, editor, *Annalen der Naturphilosophie, Bd. 14, Heft 3-4*. 1921. (Nachdr.: Suhrkamp, Frankfurt, 1963.).

[Wit53] L. Wittgenstein. *Philosophische Untersuchungen*. Blackwell, Oxford, 1953. (Nachdr.: Suhrkamp, Frankfurt, 1984.).

[Wro91] S. Wrobel. Towards a model of grounded concept formation. In *Proceedings of the 12^{th} International Joint Conference on Artificial Intelligence*, pages 712–717, 1991.

Stichwortverzeichnis

Autorenverzeichnis