



GMD –  
Forschungszentrum  
Informationstechnik  
GmbH

Franco di Primio

## Roboter cognition

Aufsetzpunkte und Probleme für eine  
biologisch-evolutionär orientierte,  
synchronisch und diachronisch  
minimierende Forschung

© GMD 1998

GMD –  
Forschungszentrum Informationstechnik GmbH  
Schloß Birlinghoven  
D-53754 Sankt Augustin  
Germany  
Telefon +49 -2241 -14 -0  
Telefax +49 -2241 -14 -2618  
<http://www.gmd.de>

In der Reihe GMD Report werden Forschungs- und Entwicklungsergebnisse aus der GMD zum wissenschaftlichen, nicht-kommerziellen Gebrauch veröffentlicht. Jegliche Inhaltsänderung des Dokuments sowie die entgeltliche Weitergabe sind verboten.

The purpose of the GMD Report is the dissemination of research work for scientific non-commercial use.

The commercial distribution of this document is prohibited, as is any modification of its content.

**Anschrift des Verfassers/Address of the author:**

Dr. Franco di Primio  
Institut für Systementwurfstechnik  
GMD – Forschungszentrum Informationstechnik GmbH  
D-53754 Sankt Augustin  
E-mail: [franco.diprimio@gmd.de](mailto:franco.diprimio@gmd.de)

ISSN 1435-2702

*All people are ignorant,  
only on different subjects.  
Will Rogers*

## **Zusammenfassung**

Dieser Beitrag besteht aus drei Teilen. Der erste Teil ist eher programmatisch und soll als übersichtartige Vororientierung zu dem im Titel plakativ angedeuteten Problemfeld dienen. Ziel ist die zugrundeliegende Forschungsmotivation und -Fragestellung sowie die einzuschlagende grobe Vorgehensweise klarzumachen. Es wird im wesentlichen argumentiert, daß eine technische Rekonstruktion kognitiver Fähigkeiten in physikalisch »verkörperter« Form, d. h. auf der Basis von Robotern, um effektiv zu sein, biologisch orientiert und minimalistisch, d. h. auf primitive Leistungen zielend, sein muß, wobei dies in systematischer (im Sinne eines Anspruchs auf naturwissenschaftliche Erkenntnis) und methodischer Hinsicht (im Sinne abgestufter, empirisch experimenteller, z.B. auch auf die Berücksichtigung kognitiver Fehlleistungen ausgerichteter Untersuchungen) relevant ist. Im zweiten Teil werden im Hinblick auf konkrete Forschungs- und Entwicklungsarbeiten mögliche Aufsetzpunkte, wie beispielsweise die Konzepte der klassischen und instrumentellen Konditionierung, dargestellt, für welche einerseits eine biologische Fundierung sichtbar ist bzw. erreichbar erscheint, und andererseits informatisch-technische Umsetzungen bereits ansatzweise vorliegen. Der letzte Teil diskutiert einige Probleme, die sich grundsätzlich im Zusammenhang mit einer Rekonstruktion kognitiver Leistungen stellen. Diese (u.a. Reflexivität, Zirkularität, Skalierbarkeit, Modularisierbarkeit) gilt es insbesondere zu berücksichtigen, damit über eine mögliche technische Wertschöpfung hinaus auch eine theoretische erzielt werden kann.

## **Schlüsselworte**

SOZIALE KOGNITION – LEIBLICHKEIT – KONDITIONIERUNG – REFLEXIVITÄT – ZIRKULARITÄT – SKALIERBARKEIT – MODULARISIERBARKEIT

## **Abstract**

This paper consists of three parts. The first one is programmatic and intends to give a rough overview of the topics mentioned in the title. The goal is to clarify the basic motivation and research questions as well as the methodology to be used. The essential thesis is that a technical reconstruction of cognitive capabilities in a physically »embodied« form, i.e. on the basis of robots, to be effective, must be biologically oriented and minimalist, i.e. aiming at primitive capabilities. This will have a systematic impact, in the sense of a (natural) scientific claim, and a methodological effect, in the sense that gradual experimental studies are required which also explicitly consider, for instance, cognitive failures. In the second part, different starting points for research and development are examined, like the concepts of classical and operant conditioning, for which a biological basis is existing or in sight and tentative computer-based implementations are already available. The last part discusses some very fundamental problems that cannot be avoided in reconstructing cognitive capabilities. These (for instance, reflexivity, circularity, scalability and modularity) must be taken into particular account if not only technically, but also theoretically valuable results are to be achieved.

## **Keywords**

SOCIAL COGNITION – EMBODIMENT – CONDITIONING – REFLEXIVITY – CIRCULARITY – SCALABILITY – MODULARITY



# 1. Vororientierung: Roboterkognition als Programmatik

## 1.1 Begriffsreflexion und -abgrenzung

Der Terminus „Roboterkognition“ enthält zwei Aspekte, die als Hauptfragen zu verstehen sind: a) Was ist überhaupt unter „Kognition“ bzw. kognitiven Fähigkeiten zu verstehen, und b) in welchem Sinne können diese auf Roboter bezogen werden, die keine biologische, sondern technische Systeme sind.

Die den ersten Aspekt betreffenden Antworten, die man in umfangreicher nicht nur rein psychologisch orientierter Literatur findet, sind sehr differenziert. Eine lesenswerte Übersicht bietet Strube im Wörterbuch der Kognitionswissenschaft [Stru 96, S. 303-317]. Varela nimmt eine gründliche Wertung verschiedener theoretischer Stadien der neueren Kognitionsforschung aus konstruktivistischer Sicht vor [Va 88]. Aus biologischer Sicht wird Kognition von manchen Autoren (vgl. z.B. [Bi 91]) auch in einem sehr weiten Sinne gesehen, nämlich so, daß sie alle Vorgänge umfaßt, die es einem Organismus erlauben, das zu erkennen, was für sein Überleben von Bedeutung ist.

Hier möchten wir nicht einmal ansatzweise den Versuch einer vergleichenden Darstellung unternehmen, sondern lediglich folgendes (allgemein wertend und sicherlich etwas pauschalierend) festhalten. Wenn man von den Arbeiten der Kognitiven Neuropsychologie (vgl. [PoRa 94] als allgemein zugängliche Einführung) absieht, die – grob gesagt – die »Arbeit« des Gehirns während verschiedener »Denkvorgänge« untersucht und dafür chemophysikalisch fundierte Meßmethoden entwickelt hat, sind die anderen Ansätze, insbesondere diejenigen der Kognitiven Psychologie meistens Systematisierungsversuche auf der Basis einer Kombination von introspektiven und experimentellen Untersuchungen, deren Wiederholbarkeit – wenn überhaupt – sich (leider) kaum im Bereich des Zufriedenstellenden bewegt. Schlüsselbegriffe wie Denken, Erkennen, Verstehen, Vorstellen, Problemlösen bleiben letztlich dem intuitiven Verständnis des Lesers überlassen, und werden eher exemplarisch dargestellt als im analytischen Sinne auf faßbare Primitiva zurückgeführt. Eine grundsätzliche Schwäche besteht weiterhin darin, daß die gelegentlich aufgestellten »mechanistischen« Systemmodelle nicht naturwissenschaftlich, sondern in erster Linie technologisch fundiert sind (ursprünglich wurde z. B. das Gehirn als eine sehr komplexe Telefonzentrale gesehen, inzwischen sucht man Analogien zu Computern). Kritiker (vgl. [Kl 88]) gehen so weit, daß diesen »techno-analogischen« Versuchen jeder Erklärungswert abgestritten wird: die moderne Psychologie formuliere die vermeintlich ablaufenden kognitiven Prozesse lediglich in eine Sprache um, die wissenschaftlicher klingt (ib. S. 25)<sup>1</sup>.

---

<sup>1</sup> Sicherlich etwas zu weit geht der Autor, wenn er gleich am Anfang seines Buches schreibt: „modern psychology is not only valueless, but actually corrosive, destroying any possibility of insight into human behaviour“ [Kl 88, S. 1].

Im Kontext der KI-Forschung, die seit Anbeginn auch auf die Realisierung von Robotern als intelligenten Systemen abzielt, sind psychologische Begriffe und Fragestellungen weitgehend positiv rezipiert und meistens bewußt systemtechnischen Ansätzen zugrundegelegt worden<sup>2</sup>. Trotz (oder vielleicht gerade wegen) vieler Vereinfachungen sind die Begrifflichkeiten allerdings nicht einheitlich. Die Präzisierungen sind jedenfalls als nicht ausreichend für eine *durchgehende* technische (re)konstruktive Umsetzung zu werten<sup>3</sup>. Die oft angestrebte kognitive Adäquatheit von (KI-)Systemen entpuppt sich in der Realisierung als Juxtaposition metaphorischen und technischen Vokabulars (*wishful naming and comments*) ohne jegliche naturwissenschaftliche Fundierung. Die Verarbeitungsmechanismen (alias Problemlösungsprozesse), die über Repräsentationen (Gedächtnisinhalte) arbeiten, hängen letztlich in der Luft. Dies wird seit einiger Zeit zunehmend bewußt als Defizit erkannt, und konkretisiert sich in intensiveren Arbeiten zum Thema »grounding« (vgl. [Har 90], [MaSm 91] und [Wr 91]). Im expliziten Zusammenhang mit Robotik ist das Attribut »kognitiv« bislang auch nur eine programmatische Richtungs- bzw. Zielangabe, und bezeichnet keine vorhandene Eigenschaft existierender Roboter [Ch 98].

Das heißt natürlich nicht, daß im Rahmen der kognitiv orientierten KI bzw. Kognitionswissenschaft keine relevanten und präzisen Ergebnisse erzielt worden seien und weiterhin werden. Fraglich ist auf der einen Seite nur, ob die auf der Erfindung des programmierbaren digitalen Rechners gestützte Annahme bzw. Analogie („Perhaps the mind stands to the brain in much the same way that the program stands to the computer.“ [JL 88, S. 8]) mit allen damit verbundenen Konsequenzen aufrechterhalten und letztlich naturwissenschaftlich fundiert werden kann. Ohne einen Bezug zur mathematischen Theorie der Berechenbarkeit, was eine (chrono)logische Voraussetzung für digitale Rechner ist, wird andererseits eine Abgrenzung dessen, was Kogniti-

---

<sup>2</sup> Das gibt der KI-Forschung »psychologisch« gefärbte Züge, die zum Teil sehr umstritten sind. Man weiß z. B. von J. McCarthy, der einer der Gründungsväter der KI ist und für ihre mathematisch-logische Fundierung wesentliche und prägende Beiträge geleistet hat, daß er seine einführende Seminare in die KI in der Regel mit der dem Sinn nach wiedergegebenen Bemerkung anfängt: AI is a branch of computer science, not psychology. Schriftlich schwächt er allerdings diese Position etwas ab: „artificial intelligence is best regarded as a branch of computer science rather than as a branch of psychology“ [McC 88, S. 308]. Damit konzediert er, daß die »psychologische« Interpretation auch eine Option ist, wenn auch nicht die beste. An der gleichen Stelle gibt er auch zu (fast mit einem spürbaren Ton des Bedauerns), daß „artificial intelligence cannot avoid philosophy“ [McC 88, S. 305]. Die KI operiert eben nicht in einem geschlossenem Denkraum.

<sup>3</sup> Man beachte: wir sprechen von technischer (Re)Konstruktion, die wohl zu unterscheiden ist von Reproduktion (vgl. [dP 93, S. 14ff]). Der Terminus kann auch im Sinne von »reverse engineering« verstanden werden: „Reverse engineering is a technique of reasoning that works like this. You are an engineer, confronted with an artifact you have found and don't understand. You make the working assumption that it was designed for some purpose. You dissect and analyze the object with a view to working out what problem it would be good at solving: 'If I had wanted to make a machine to do so-and-so, would I have made it like this? Or is the object better explained as a machine designed to do such-and-such?'“ [Da 95, S. 103].

on im Unterschied zu Nicht-Kognition ausmacht, u.E. sicherlich nicht möglich sein (vgl. z.B. [Pen 87]).

Soweit läßt sich zum Problemfeld »Kognition« aus KI-Sicht trotz vieler Forschungsanstrengungen und gelegentlichen Vereinheitlichungsansätzen (vgl. [Ne 90]) jedenfalls gerechtfertigt feststellen, daß es kaum umfassende Theorieversuche geschweige denn »geschlossene« Theorien gibt. Der Sprachgebrauch ist schwankend; Kategorisierungs- (eigentlich mehr lexematische Gruppierungs)versuche in Lehrbüchern sind schwer verständlich und bringen keine Klarheit, sie machen ja nur klar, daß Vieles noch unklar ist [Mü 98].

## **1.2 Vorgehensweise**

Unsere Vorgehensweise kann aus verschiedenen Blickwinkeln beschrieben werden. Grundsätzlich ist sie sowohl als theoretisch als auch als operational, d. h. experimentell und systemtechnisch orientiert, bzw. offen zu verstehen. Offenheit ist im Sinne einer progressiven, selektiv um sich greifenden Assimilation (des interdisziplinären Forschungsumfeldes) gemeint.

### **1.2.1 Kognitionswissenschaftlicher Forschungskontext**

Eine evaluierende Untersuchung (beinah) geschlossener Theorien (z. B. Minsky's Society of Mind [Mi 86]) soll zu einer Explizierung ihrer Grundannahmen führen. In dieser Richtung bieten sich folgende schematische Einstiegsfragestellungen an:

- Werden die Begriffe »Kognition« und »Intelligenz« überhaupt voneinander abgegrenzt? Wozu ist Kognition gut? Ist Kognition eine systemimmanente Eigenschaft oder (nur) eine Beobachtungskategorie?
- Wird Kognition als eine »einheitliche« Fakultät aufgefaßt? Welche Primitiva (Bausteine bzw. Mechanismen) werden identifiziert, und wie sehen die (statischen bzw. dynamischen) Kompositionsregeln aus?
- Wird Kognition als ein »Wissen« (inneres Phänomen, Aufbau von Konzepten und Vorstellungen, Erkennen) und/oder »Können« (externes Phänomen, umgebungsverankert und -verändernd) verstanden? Was ist (woran erkennt man) kognitives »Verhalten«?
- Sofern (technische) Rekonstruktionen bzw. Modelle aufgestellt werden: was ist in solchen Modellen technologische Analogie und was naturwissenschaftliche Basis?

Der Vergleich mehrerer Theorieansätzen wirft zusätzliche, sehr weitreichende Fragen auf. Diese betreffen in erster Linie das Verhältnis von Kognition zur Informationsverarbeitung (sind kognitive Prozesse identisch mit Berechnung?) und Kybernetik. Der Stellenwert und die Tragweite der Physical Symbol System Hypothesis, die besagt, daß die formale Manipulation von Symbolen ein sowohl notwendiger als auch hinreichender Mechanismus zur Realisierung von allgemeinem intelligentem Verhalten ist [Ne 80], ist im Zusammenhang mit verteilten, parallelen Ansätzen wie Konnektionismus

bzw. künstliche neuronale Netze ([RuMC 86], [HaBu 90], [Pal 97]) und dynamischen Systemen ([Cl 97a], [Hei 97], [Jae 96]) zu diskutieren. Ferner ist der Beitrag der konstruktivistischen Sicht ([Gla 97], [RiVe 91]), sowie die mögliche theoretische und/oder technologische Auswirkung neuer Computer-Architekturen wie dem Quantencomputer ([Pr 98], [Ri 97], [BeDiV 95]) zu reflektieren<sup>4</sup>. Es versteht sich, daß nicht nur eine Klärung, sondern allein das Erfassen solcher Zusammenhänge eine nur interdisziplinär zu bewältigende und sicherlich lange andauernde Aufgabe ist.

## 1.2.2 Defizite gegenwärtiger Theorien

Mögliche Defizite gegenwärtiger Theorien lassen sich beim Versuch einer Beantwortung folgender Fragen festmachen (vgl. auch [dPM 94]):

- In welchem Sinne ist Kognition ein »Individual«-Phänomen?
  - Innerer Bezug: Verhältnis von Kognition zu anderen »Bereichen« wie Emotion, Motivation, Wahrnehmung, Gedächtnis, Bewußtsein.
  - Bio-physikalischer (äußerer) Bezug: Verhältnis von Kognition zur »Körper«-Problematik (Leiblichkeit): Thematisierung des Problems, was ist in einem (kognitiven) System Innen, was ist Außen, und welche ist die Rolle des »Randes« (Sensoren, Effektoren, Gebrauch von Werkzeugen).
- In welchem Sinne ist Kognition ein »interaktives« Phänomen?
  - Sozialer Bezug: Ist Kognition eine Eigenschaft, die unabhängig vom »sozialen« Kontext verstanden werden kann, in welchem sie zum Ausdruck kommt?
  - Wie ist das Verhältnis von Selbst- und Fremdkognition (Wahrnehmung & Erkennen von sich selbst und von Artverwandten).
  - Was ist das Verhältnis von Kognition zu Kommunikation (Sprache und Verhaltenskoordination)?

---

<sup>4</sup> „Whereas ordinary classical computers process information encoded in bits, a quantum computer processes information encoded in quantum states ... A quantum computer exploits a kind of massive parallelism that can never be approached by any conceivable conventional digital computer.“ [Pr 98, S. 631]. Die Brisanz (und Problematik) dieser Entwicklung, die in kognitionswissenschaftlichen Kreisen bislang kaum rezipiert worden ist, zeigt sich in Folgendem: Roger Penrose hat im Zusammenhang mit menschlicher Kognition die Frage aufgeworfen, ob bewußtes Denken (conscious thinking) auf Berechnungen im Sinne einer Turing Maschine reduzierbar ist. Seine Antwort (wobei er seine Argumentation auf Gödel's Theorem stützt) ist negativ [Pen 87]. Weiterhin stellt er die Hypothese auf, daß für ein Verständnis menschlichen Denkens auf physikalischer Ebene die Quantenmechanik zu berücksichtigen sei: „It seems to be the prevailing view that classical physics should be adequate for describing behaviour on the scale relevant to a human brain. Nevertheless, it could be the case that *quantum mechanics* is actually needed.“ [ib., S. 271]. Damit würde man aber nach seiner Meinung letztlich bei einer nicht-algorithmischen Physik enden [ib., S. 274]. Es ist (vielleicht nicht nur für uns) unklar, wie diese Überlegungen gerade mit dem Fortschreiten in Richtung Quantencomputer zusammenhängen, vor allem angesichts der Tatsache, daß dafür bereits *Algorithmen* (z.B. für die Faktorisierung von Zahlen) vorgeschlagen worden sind [Pr 98, S. 631].



Aus den Defiziten leitet sich nicht nur unsere Motivation, sondern auch die konkrete Vorgehensweise und Zielsetzung ab.

### **1.3 Zielsetzung: biologisch fundierte soziale Kognition**

Was die Zielvision angeht, so möchten wir (selbstverständlich im Sinne einer erweiterten, kooperativen Forschungsanstrengung) als Antwort auf die Frage des Verhältnisses von Kognition zur Körper-Problematik idealiter konkrete technische, d. h. hard-, softwaremäßig sowie mechatronisch verankerte und biosoziologisch fundierte Lösungen erreichen. Diese verankerte Form der (gruppenorientierten, kollektiven bzw. sozialen) Kognition, die wir abkürzend als *social cognition grounding* bezeichnen wollen, ist ein diesen von (vielen) früheren KI-Forschungsansätzen unterscheidendes Merkmal<sup>5</sup>. Damit wird dem Grundtenor der *social intelligence hypothesis* [Jo 66] Rechnung getragen, die in einer besonderen Ausprägung besagt, „that intelligence in biological systems has been evolved out of social needs in groups of conspecifics and forecasting the very next movements of each another by a kind of inner rehearsal“ [Ch 98, S. 497], und der Fokus der Aufmerksamkeit auf die Phänomene der Kooperation bzw. Konkurrenz (vgl. z.B. [Er 98] und [CaKa 95]) und Kommunikation gerichtet.

Das Thema *soziale Kognition* wird aus biologischer Sicht in der Verhaltensforschung schon seit geraumer Zeit untersucht und mit einem besonderen Stellenwert versehen. D. Bischof-Köhler sieht z. B. darin „den qualitativen Sprung vom evolutionären Niveau der niederen Primaten zu den Anthropoiden“ [Bi 91, S. 146]. Dieselbe Autorin weist auch insbesondere auf die entscheidenden Aspekte der Kooperation und des Teilens von Nahrung hin, die den Aufbau von inneren Modellen und *mentalen Bühnen* zum *Probhandeln* mitbedingen bzw. voraussetzen. „Die Jagd auf Großwild erfordert [...] ein anspruchsvolleres Vorgehen. Man muß die Intention des Jagdkumpans verstehen, sie aufgreifen, dann aber nicht das Gleiche tun wie er, sondern die eigenen Aktivitäten komplementär auf sein Vorgehen abstimmen, damit ein Ziel gemeinsam erreicht werden kann.“ [Bi 91, S. 152]. „Analysiert man das Jagdverhalten der Schimpansen [...], dann zeigen sie in der Tat eine gewisse »Einsicht« in das Wesen ihrer Beutetiere, etwa wenn einige Tiere den Fluchtweg des Opfers voraussehen, welches ihr Kumpan jagt, und sich so gruppieren, daß sie es abfangen können.“ [Bi 91, S. 166 f]. Das Verstehen von Intentionen und – psychologisch gesprochen – eine »Rollen«- und »Perspektivübernahme« sind ohne interne Modellierung und Phantasietätigkeit (Imagination), die im Gegensatz zum einfachen Erinnern erfordert, „daß Vorstellungs-

---

<sup>5</sup> Neuere Forschungsansätze, die auch auf die Rekonstruktion von »sozialer Intelligenz« zielen, erheben (ganz im Sinne der symbolverarbeiteten KI) keinen grounding-Anspruch, bzw. setzen dieses Problem als gelöst voraus. So schlägt Worden ein „working computational model of social intelligence“ vor, deren Hauptbestandteil ein „Social Intelligence Module“ ist, welcher „continually receives precategoryed, symbolic inputs from other cognitive subsystems such as the visual system.“ [Wo 96, S. 579, 582, 595].

inhalte *aktiv verändert* und in neue Zusammenhänge gebracht werden können“, nicht möglich [Bi 91, S. 166]. Das individuelle, innere Durchspielen von Lösungen als Mittel zur Bewältigung schwieriger Situationen [Kli 93, S. 170] setzt die Fähigkeiten zum Rückschauen und Vorausschauen voraus. Diese sind aber letztlich höchstwahrscheinlich zwei Seiten derselben »sozial-kognitiven Medaille«: Konspezifika (Artverwandte) und Kumpane fungieren sozusagen als »Spiegel« (und als Weg zum Verständnis) von einem selbst.

Unsere in diese (im Sinne des rekonstruktiven Anspruchs *ultimative*) Zielrichtung leitende Basis-Arbeitshypothese ist, daß Kognition die Entkoppelung von Reiz-Reaktionsvorgängen voraussetzt, und daß sie eng mit der Problematik der Gedächtnisorganisation zusammenhängt. Kognitive Systeme sind dadurch gekennzeichnet, daß in ihnen eine indirekte Kopplung zwischen Reiz und Reaktion besteht, oder, positiv umformuliert, daß bei ihnen (eben kognitive) Prozesse modifizierend zwischen den Reiz und die Reaktion (das Verhalten des Systems) geschaltet sind ([Mü 98], [Stru 96, S. 304]). Die Tatsache, daß auf der Ebene der vermittelnden Prozesse wieder das Attribut kognitiv verwendet wird, verdeutlicht, daß die formulierte Eigenschaft (Entkoppelung) zwar eine notwendige Bedingung darstellt, daß sie aber nicht hinreichend ist. Reimann und Wolff [ReWo 98], die sich auf molekularbiologische Daten stützen, sprechen auch von vermittelnden (modulierenden) Prozessen, die homeostatisch gerichtete Adaption bewirken (ib., S. 7). Das ist allerdings in einem grundsätzlichen Sinn gemeint: Sie sehen darin eine (vielleicht die wesentliche) Voraussetzung für die Emergenz von Kognition, weil durch die Speicherung und Integration von Signalen über die Zeit eine Vorausschau-Fähigkeit entsteht: das „adapting system functions as *predictor*“ (ib., S. 8).

Eine weitere Grundannahme ist, daß »Kognition« ein im doppelten Sinne verhaltensrelevantes System von Eigenschaften ist: sie dient der (im o. e. Sinne *sozialen*) Verhaltenskoordination und ist letztlich beobachtbar in ihren Auswirkungen.

## 1.4 Minimalistische Methodik

Methodisch allgemein möchten wir so weit wie nur möglich operationalisierbare Behauptungen aufstellen. Mit operationalisierbar meinen wir, daß es letztlich darum geht, Antworten auf konkret überprüfbare Fragen zu finden: Was soll ein »kognitiver« Roboter können, was ein »einfacher« Roboter nicht kann? Wie und wodurch kann der erste das, was der andere nicht kann? Im Idealfall sollte man zeigen können, wie eine kognitive (Verhaltens)Leistung durch (sukzessive) Entfernung bzw. Veränderung eines oder mehrerer Bestandteile der Theorie bzw. (der Bausteine bzw. Mechanismen) eines Systems »degradiert« bzw. gänzlich zusammenbricht. Damit sind auch *kognitive Fehlleistungen* explizit im Fokus der Aufmerksamkeit. Man beachte, daß wir in dieser Hinsicht völlig im Einklang mit einem Grundsatz biologischer naturwissenschaftlicher Vorgehensweise sind: „In Biologenkreisen gibt es ein Axiom, das

besagt, wenn man verstehen wolle, wie etwas funktioniert, so müsse man es im gestörten Zustand untersuchen.“ [Ga 89, S. 43]. *Normalitätsmodelle* für komplexe Systeme (wie Lebewesen) lassen sich am besten inkrementell aufstellen, indem man Fehlzustände (und -modelle) untersucht.

Weiterhin sind wir der Meinung, daß man dem Phänomen Kognition nur dann im Sinne eines erfolgsversprechenden technisch rekonstruktiven Ansatzes beikommen kann, wenn man minimalistisch vorgeht. Das heißt, man stellt minimale Randbedingungen für das Vorhandensein bzw. die Realisierung von Kognition auf. In diesem Sinne kann z.B. »Verhalten« zu »Bewegung(en)« minimiert werden. Daraus ergibt sich eine Konkretisierung der weiter oben formulierten Frage: Was unterscheidet eine kognitive von einer nicht-kognitiven Bewegung? Ist eine nicht-zufällige, zielgerichtete Bewegung bereits als kognitiv zu bewerten?

Ermutigende erste Ergebnisse im Sinne dieser Methodik sind von V. Braitenberg erzielt worden. Seine Arbeiten [BrSc 93] zeigen, daß komplex interpretierbares Verhalten (Einstellungen wie z.B. Ab- und Zuneigung und Aggressivität) auf sehr einfache Mechanismen und Verknüpfungsmuster zurückführbar ist. Man beachte allerdings, daß in diesem Kontext, genauso wie in unserem, keine Minimierung im Sinne der Physik angestrebt wird, die mit *idealisierten* Begriffen wie eben z. B. »Bewegung ohne Reibung« operiert. Eine biologisch orientierte Minimierung der Bewegung stellt eine offene Frage dar, es ist sogar zu klären, ob der Rückgriff auf elementare Organismen überhaupt zu einer Vereinfachung führt. Sind z.B. amöboide Zellbewegungen, die unter ständiger Veränderung der Zellform stattfinden, einfacher (im Hinblick auf eine technisch-rekonstruktive mechatronische Verwertbarkeit) als Wurm- oder Schlangenbewegungen?

Wir möchten das Ziel der Realisierung *biologisch* fundierter Roboter (Artefakte) im Auge behalten, glauben allerdings, daß die Machbarkeit nur durch evolutionär orientierte Komplexitätsreduktion überprüft werden kann. Dadurch ergeben sich Differenzierungsmöglichkeiten in dem „nach unten“ gerichteten Abgrenzungsversuch [St 96, S. 304]. Eine effektive Reduktion und (letztlich auch methodisch relevante) Minimierung soll *diachronisch und synchronisch* erfolgen.

Synchronisch bedeutet, daß man gegenwärtige Lebensformen betrachtet, die im Sinne der (hypothetisch angenommenen) kognitiven Eigenschaften eine primitive oder gar keine Leistungsfähigkeit aufweisen<sup>6</sup>. Strube führt als Beispiele folgende Fragen auf [St 96, S. 304]: Ist die Fliege ein kognitives System; ist die geschickte Steuerung ihres Fluges eine kognitive Leistung? Aus biologischer Sicht liegen im Sinne vergleichen-

---

<sup>6</sup> Beer und andere nehmen auch eine biologische (neuro-ethologische), in unserem Sinne synchronisch minimierende Perspektive ein, sie untersuchen allerdings primitives (Insekten-) Verhalten mit Blick auf das Problem der Adaptivität, ohne die Frage der Kognition explizit zu stellen („our work has focused on the behavior of simpler natural animals“ [BCS 91, S. 170]).

der Studien der Intelligenz verschiedene Ergebnisse vor. Eine nachdenklich stimmende kritische Wertung liefert E. Macphail, der letztlich – von der menschlichen Art abgesehen – zumindest in Bezug auf Wirbeltiere für eine „null hypothesis“ plädiert: „there are in fact no differences – qualitative or quantitative – among the intellects of non-human vertebrates.“ [Mac 87, S. 180].

Ist die synchronische Betrachtung eher »geographisch« orientiert, so wird die diachronische Minimierung dagegen entlang der Zeitachse (also »historisch«) vorgenommen, d. h. dadurch, daß frühere Entwicklungsstadien von Lebewesen im Sinne der Fragestellung untersucht werden. Wann ist Kognition in der Bioentwicklung aufgetaucht, woran kann das festgemacht werden? In diesem Sinne ist das Stichwort »evolutionär« im Beitragstitel zu verstehen. Am schwierigsten dürfte in diesem Zusammenhang sein, einen sowohl effektiven als auch effizienten Zugang zu einer technischen Umsetzung dessen zu finden, was manche Biologen den evolutionären *Schneeballeffekt* nennen:

The snowballing effect is summed up in an informal principle formulated by Jacob, a cell geneticist. It is just this: 'Simpler objects are more dependent on (physical) constraints than on history. As complexity increases, history plays the greater part.' [...] Jacob compares evolution to a tinkerer who must use whatever is immediately at his disposal to achieve some goal. This case contrasts with that of an engineer who, within certain limits, decides on the appropriate materials and the design, gathers them, and then carries out the project. [...] The point, then, is that what the tinkerer produces is heavily dependent on his historical situation in a way in which the engineer's product is not. [Cl 97, S. 70].

Ein solcher Zugang sollte eine *Inkrementalität* bei der Rekonstruktion von Kognition ermöglichen und letztlich eine Beantwortung der Frage erlauben, ob überhaupt und in welcher Abstufung Kognition als »graduelles« Phänomen zu verstehen ist.

Man beachte schließlich, daß die phylogenetische Betrachtung eine Beschäftigung mit Kollektivphänomenen erzwingt, d. h., es geht letztlich um die Entstehung von Kognition in Populationen von Lebewesen (und Robotern). Eine biologische orientierte, auf Minimalitätsannahmen basierende Vorgehensweise erfordert (im maximalen Fall) eine Erklärung der (selbstreproduktiven) Bedingungen bzw. Mechanismen, unter denen selektive Übergänge zwischen (kognitiv komplexer werdenden) Populationen erfolgen.

Wissenschaftliche Aktivitäten mit einer solchen Zielsetzung sind im Ansatz bereits vorhanden. So sprechen Dautenhahn und Christaller von „very first steps towards artificial social intelligence for robots“ [DauCh 95, S. 16]. Sie fokussieren allerdings individualisierte Robotergesellschaften, d. h. solche, in welchen die Individuen sich gegenseitig (er)kennen, und setzen damit auf eine (wahrscheinlich) sehr hohe phylogenetische Entwicklungsstufe auf. Fukuda und Ueyama, die den Begriff »cellular robotics« eingeführt haben, forschen auch bereits seit einiger Zeit in die »soziale« Rich-

tung, scheinen aber mit der Miniaturisierung der Robotersysteme hauptsächlich eine Basis zu suchen (Roboterzellen), um, biologisch gesprochen, die onto- und morphogenetische Entwicklung nachzumachen [FuUe 94]. Mit annähernd ähnlicher Motivation spricht Connell von „Minimalist Mobile Robotics“ und stellt eine Roboterarchitektur dar, die aus einer »Kolonie« von autonomen »agents« besteht, wobei ein bestimmter Typ von Schnecke als biologisches »Vorbild« dient [Co 90]. Agah und Bekey untersuchen dagegen explizit mit Hilfe simulierter Roboterkolonien den Zusammenhang zwischen phylogenetischem und ontogenetischem Lernen [AgBe 97]. Diese Beziehung dürfte die ultimative Forschungsfrage in Bezug auf die Entwicklung von Kognition sein.

## 1.5 Forschungsansatz im Überblick

Zwischenzusammenfassend läßt sich unser Ansatz wie folgt darstellen:

- Konzentration auf primitives (räumliches) Verhalten von (Gruppen von) einfachen Artefakten bzw. (gegenwärtiger und/oder früherer) Lebewesen
- Identifikation (primitiver) Kognitionsleistungen und Klärung der biologischen Grundlagen für mögliche (re)konstruktive Annahmen
- Aufstellung differenzierender minimaler Szenarien:  
Rolle von »Gedächtnis« und »Verarbeitung« von »Erinnerungen« (Rückschau, Lernen, Vorausschau,) beim »Verhalten« und anderen Aspekten (wie z.B. Motivation)
- Wie sieht eine (selektionsbasierte) Progression der (kognitiven und sonstigen) Fähigkeiten der Artefakten (Robotern) bzw. Wesen aus?  
Wie sieht eine dem (evolutionsbasierten Gedanken) Rechnung tragende Folge von (minimalen) Szenarien aus?
- Durchführung von Experimenten durch Simulation und/oder mit (Mini-) Robotern

## 2. Aufsetzpunkte für eine biologische Fundierung

Welche konkrete Einstiegsmöglichkeiten sind im Rahmen der vorgestellten allgemeinen Auffassung und Vorgehensweise gegeben, die eine minimierende, biologisch relevante Orientierung, wenn nicht letztlich Fundierung der Rekonstruktion kognitiver Leistungen, erlauben?

Wir identifizieren zunächst mehrere Bereiche, die mit sogenannter Konditionierung (auch »assoziatives Lernen« genannt), (Wieder)Erkennen und (einsichtigem) Lernen zu tun haben. Deren gemeinsame Basis könnte die auf primitive Verstärkungs- und Abschwächungsmechanismen beruhende, von Donald H. Hebb neuropsychologisch interpretierte Idee der *Assoziationen* sein [PoRa 94, S. 6f], die Hassenstein als *Wir-*

*kungsquanten* der Intelligenz bezeichnet [Ha 92]<sup>7</sup>. Schließlich zeigen wir, wie durch eine körper- und verhaltensbezogene Deutung von Konzepten auch höhere kognitive Leistungen vereinfachend in Angriff genommen werden könnten. Inwiefern all diese Aspekte synchronisch verteilt sind bzw. diachronisch mehreren Ebenen der »kognitiven Anpassung« zugeordnet werden können, ist eine offene Frage<sup>8</sup>.

## 2.1 Klassische Konditionierung

Die von Pawlow etablierte Forschung zur Konditionierung hat Ergebnisse geliefert [Wa 82], die u.E. als Basis für jeden rekonstruktiven Versuch in einer kognitiv orientierten Robotik dienen dürften, die neurobiologisch fundiert sein will<sup>9</sup>. Die Unterscheidung zwischen primitiven Reaktionen (unbedingte Reflexe) und Ketten von aufeinanderfolgenden einfachen Reaktionen (Instinkte) mit dem Reflexbogen als Strukturprinzip bietet eine Grundlage für zwei entscheidende Aspekte der Architektur eines Roboters. In biomorphischer Ausdrucksweise: die *Unverletzbarkeit* und *Überlebensfähigkeit*. Nach Pawlows Meinung sind die unbedingten Reflexe und Instinkte die Grundvoraussetzung für die Unverletzbarkeit eines Organismus in der Natur. Die auf den unbedingten aufbauenden bedingten Reflexe, mit denen Reize, die unbedingten Reizen vorausgehen, zu Stellvertretern (Signalen) der unbedingten Reize und somit zu (vor- und deshalb auch rechtzeitigen) Auslösern der unbedingten Reaktionen werden, erlauben den Aufbau komplizierter Leistungen im Sinne der Überlebensfähigkeit eines Organismus.

Die Mächtigkeit und theoretische Tragweite bedingter Reflexe ist erheblich, weil bedingte Reize sich sozusagen vermittelnd, letztlich im Sinne einer stellvertretenden (*repräsentierenden*) Funktion, auswirken:

Dadurch, daß der Organismus fähig wird, nicht nur auf unbedingte Reize zu reagieren, sondern auch auf die *vorhergehenden* Signale, ist er auf den *Zeitpunkt* und die *Art* des eintretenden Ereignisses *vorbereitet*. Zwischen die Antwortreaktion und die Wirkung ihres spezifischen Auslösers, des unbedingten Reizes, wird gewissermaßen ein zusätzliches Glied geschaltet, das Signal, welches das eintretende Ereignis repräsentiert. [Wa 82, S. 14]

Der mit dem bedingten Reiz verbundene Wechsel der Sinnesmodalität (der bedingte Reiz kann z. B. akustisch oder optisch, während der unbedingte Reiz eine Kontaktwahrnehmung sein kann) kann lebenswichtige Leistungen im Sinne von »entfernten«

---

<sup>7</sup> Beispiele für kognitionswissenschaftliche Arbeiten, die auf einfache Lernprozesse der Konditionierung aufsetzen, liefern [Mur 96] und [Ver 98]. Poon und Shah stellen in [PoSh 98] eine (künstliche) neuronale Architektur dar, deren synaptische Lernregel auf Hebb's Idee basiert.

<sup>8</sup> N. Bischof unterscheidet drei stammesgeschichtlich relevante Ebenen [Bi 91, S. 162 ff]: die Ebene der instinktiven Verhaltensanpassung, die Ebene der Phantasietätigkeit bei den Menschenaffen und schließlich das Niveau der spezifischen menschlichen Kognitionsleistungen. Diese grobe Unterscheidung ist u. E. nicht besonders hilfreich.

<sup>9</sup> In [MiMa 97] wird ein auf *künstlichen* neuronalen Netzen basierendes Modell der klassischen Konditionierung dargestellt.

Reaktionen ermöglichen (Pawlow nannte bedingte Reflexe in dieser Hinsicht auch „Reflexe auf Distanz“).

Einen systematischen Überblick über den neueren Stand der Forschung zur klassischen Konditionierung gibt Menzel in [DMS 96, S. 502 ff]. Der Autor unterstreicht die kognitiven Aspekte der klassischen Konditionierung, so insbesondere, daß zeitliche Kontiguität allein nicht hinreichend für eine erfolgreiche Konditionierung ist. Es kommt auch auf die »informationelle Beziehung« (Regelmäßigkeit der Paarung) zwischen bedingtem und unbedingtem Reiz und die Vorhersagbarkeit des unbedingten Reizes an.

Man beachte, daß sofern der Begriff »Reflex« im Sinne Pawlows als neuronaler Mechanismus verstanden wird, er eine (synchronische und diachronische) Schranke für die Suche nach biologischer Fundierung für Kognition abgibt. Wenn kognitive Leistungen auf Reflexen bauen, so darf man für Roboter cognition keine biologischen Modelle in existierenden oder vergangenen Lebewesen zu finden hoffen, die nicht (klassisch) konditionierbar sind. Sind (beispielsweise) Einzeller im Pawlowschen Sinne konditionierbar? Ist die Antwort negativ, so stellt sich die aus unserer Sicht im Sinne einer *präkognitiv* relevanten Forschung komplementäre und interessante Frage, wie aus dem Zusammenkommen von Einzellern sich Spezialisierungen in Richtung Nervenzellen gebildet haben, und wie aus diesen Reflexbögen entstanden sind<sup>10</sup>.

## 2.2 Instrumentelle Konditionierung

Die instrumentelle (auch operante genannt) Konditionierung kann als Gegenstück zur klassischen betrachtet werden [ApDö 80, S. 108ff]. Während bei der klassischen Konditionierung ein neuer Reiz (der bedingte) für eine vorhandene Reaktion gelernt wird, wird bei der instrumentellen eine neue Reaktion (Verhaltensweise bzw. motorisches Muster) entdeckt und beibehalten. Typische Beispiele sind Experimente in sogenannten Problem-Käfigen [DMS 96, S. 507 ff]: eine hungrige Katze (als Beispiel) versucht aus dem Käfig zu entkommen, indem sie lange um sich herum beißt und kratzt, bis sie zufällig den richtigen Öffnungsmechanismus erwischt. Bei Wiederholungen des Experiments gelingt dies dem Tier immer schneller, bis es den Hebel betätigt, sobald es in den Käfig gesetzt wird. So »lernt« das Tier durch »Versuch und Irrtum« (operante Aktivität: Ausprobieren) eine neue Verhaltensweise (eine erfahrungsbedingte Aktion im Unterschied zu vorgegebenen, unbedingten Reaktionen). Interessant in diesem Zusammenhang ist die Tatsache, daß die auf zufälligem Vorgehen basierenden Lernkurven nicht von den Lernkurven von Tieren (Katzen, Hunde und Hühnchen) zu unterscheiden sind, denen die Möglichkeit gegeben wird, andere Tiere beim

---

<sup>10</sup> Daß ein (auch nur) minimales Nervensystem hinreichend für die Herausbildung von Verbindungen zwischen zwei Reizen ist, scheint unbestreitbar zu sein. Nachgewiesen ist das z.B. für einen nur einige Millimeter langen Wurm (*C. elegans*), der einen Körper aus weniger als 1000 Zellen besitzt, wobei davon 302 Nervenzellen sind [BCS 91, S. 173].

Finden der Lösung zu beobachten, bzw. im Vorlauf sogar »gezeigt« wird (z. B. durch Auflegen der Pfote auf den richtigen Hebel), wie die Lösung gefunden werden kann [Th 98, S. 45 ff]. Beobachtung und Imitation haben also für (solche) Tiere anscheinend keinen oder nur einen sehr geringen operationalen Stellenwert.

Aus der Sicht unserer Fragestellung sind diese Ergebnisse von Bedeutung, weil sie erlauben, Stadien der Problemlösung und des Lernens zu differenzieren. Das zufällige Suchen ist auch für höhere kognitive Wesen eine Strategie, die als ultima ratio angewendet werden kann, d. h., wenn es nicht möglich ist, den »Problemraum« zu strukturieren.

Die instrumentelle Konditionierung wird in letzter Zeit in der KI unter der Bezeichnung *reinforcement learning* („the problem faced by an agent that learns behavior through trial-and-error interactions with a dynamic environment“) verstärkt untersucht [KaLi 96]. Dabei wird explizit Bezug auf die weiter oben erwähnten (inzwischen 100 Jahre alten) Arbeiten von Thorndike genommen, insbesondere auf seine „Law of Effect“, d. h. „the idea that actions followed by good or bad outcomes have their tendency to be re-selected altered accordingly“ [SuBa 98, § 1.6]. Die fundamentale Bedeutung dieses »Gesetzes« stellen Sutton und Barto wie folgt dar:

The Law of Effect includes the two most important aspects of what we mean by trial-and-error learning. First, it is selectional rather than instructional in the sense of Monod, meaning that it involves trying alternatives and selecting among them by comparing their consequences. Second, it is associative, meaning that the alternatives found by selection are associated with particular situations. Natural selection in evolution is a prime example of a selectional process, but it is not associative. Supervised learning is associative, but not selectional. It is the combination of these two that is essential to the Law of Effect and to trial-and-error learning. Another way of saying this is that the Law of Effect is an elementary way of combining search and memory: search in the form of trying and selecting among many actions in each situation, and memory in the form of remembering what actions worked best, associating them with the situations in which they were best. Combining search and memory in this way is essential to reinforcement learning. [SuBa 98, § 1.6]

Daraus ist ersichtlich, daß diese (»psycho-biologisch« fundierte) Forschungsrichtung relevante Beiträge im Hinblick auf eine Lösung des bereits angesprochenen (vgl. § 1.4), noch höchst klärungsbedürftigen Verhältnisses zwischen phylogenetischem und ontogenetischem Lernen liefern kann.

## **2.3 (Wieder)Erkennen: Erfassen von Identität**

Eine möglicherweise unterhalb des Mechanismus der Assoziation liegende und für jedes intelligente Verhalten erforderliche Basis-Erkenntnisleistung ist die Erfassung von Identität. »Identisch sein« bedeutet, daß zwei Phänomene, die raum-zeitlich



getrennt wahrgenommen werden, als eine Einheit erscheinen. Wegen der zeitüberbrückenden Funktion spricht man von »diachroner« Identität:

Diachrone Identität ist erforderlich, um Dinge als etwas wiederzuerkennen, das einem zuvor begegnet ist, bzw. zu erwarten, daß sie wiederauftauchen, wenn sie verschwunden waren. Lebewesen, die nicht über diese Kategorie verfügen, wie wahrscheinlich niedere Insekten oder auch Amphibien, leben in einer Welt ständig wechselnder Bilder. Selbst noch für die Schlange, die eine Maus verfolgt, hört diese auf zu existieren, sobald es ihr gelingt, in ihr Loch zu flüchten. Die meisten Säugetiere dagegen können diachrone Identität wahrnehmen. [...] eine Katze wird sich geduldig vor dem Loch auf die Lauer legen, in das sich die Maus geflüchtet hat. [Bi 91, S. 163f]

Mit der Annahme dieser primitiven kognitiven Leistung ergibt sich eine (weitere) Abgrenzungsmöglichkeit in der Suche nach biologischen Modellen für Roboterkognition. Eine Differenzierung in der Fähigkeit zum Erfassen von Identität ließe sich im Verhältnis zu anderen Wahrnehmungsleistungen vornehmen. Ist die Erfassung verschiedener Ausprägungen eines (un)bedingten Reizes keine identitätserkennende Leistung? Ist ein Lebewesen (wie die Schlange), das nicht zum Erfassen von Identität im oben erläuterten Sinne fähig ist, überhaupt (klassisch) konditionierbar? Wie hängt das Erfassen von Unterschieden mit dem Erfassen von Identität zusammen?

Auf die Grundsätzlichkeit dieser Fragestellung weist auch S. Harnad hin, der sich im Zusammenhang mit dem *symbol grounding* Problem mit dem Verhältnis von *identification*, *discrimination* und *categorical perception* beschäftigt. Für ihn gilt, daß die Fähigkeit zum Identifizieren unabhängig ist von der Fähigkeit zur Diskriminierung (Erfassen von Unterschieden): „discrimination is independent of identification“. Diese zuerst etwas unverständlich erscheinende Position wird nachvollziehbar, wenn man die Definitionen betrachtet, die der Autor gibt:

Discrimination is a relative judgment, based on our capacity to tell things apart and discern their degree of similarity. To be able to identify is to be able to assign a unique (usually arbitrary) response – a "name" – to a class of inputs, treating them all as equivalent or invariant in some respect. Identification is an absolute judgment, based on our capacity to tell whether or not a given input is a member of a particular category. [Har 90, S. 341]

Nach dieser Lesart ist aus unserer Sicht die (kognitive?) Fähigkeit entscheidend, welche im *Erfassen von Graden der Ähnlichkeit* von Objekten besteht. Ohne diese ist wohl weder Diskriminierung noch Identifizierung (im Sinne von Harnad) möglich.

## 2.4 Lernen durch Einsicht

Einsichtiges Lernen steht im Gegensatz zum Lernen auf der Basis des »blinden« Versuchs-und-Irrtums-Verhaltens der instrumentellen Konditionierung. Einsicht liegt vor, wenn eine Verhaltensänderung auf (spontanes) Erkennen der Struktur einer gegebenen Aufgabe basiert [ArEyMe 97, S. 1243].

Klassisch sind in dieser Hinsicht die Versuche von W. Köhler mit Schimpansen, die, um an eine hochgehängte bzw. außerhalb des Geheges ausgelegte Banane zu kommen, herumliegende Kisten aufeinander stellten, um sie als »Leiter« zu benutzen, bzw. Stöcke zusammensteckten, um die Frucht herbei zu angeln. Die Tiere gingen nicht zufällig vor, sondern ließen – ruhig sitzend – ihren Blick zwischen Bananen und Kisten bzw. Stöcken hin- und herwandern, um dann plötzlich aufzuspringen und die Lösung zu produzieren, als wäre ihnen diese durch »Überlegung« eingefallen [Bi 91, S. 154f].

B. Hassenstein bringt in dieser Richtung auch einige Verhaltensbeispiele aus dem nicht-Primaten-Bereich, die er an der Grenze zwischen »bloßem Lernen« (d. h. Konditionierung) und »Intelligenz« lokalisiert sieht. Er spricht von Neuverknüpfungen aus innerem Anlaß (»freie« Assoziationen), korrigierten Versäumnissen und Vergleich von Engramm und Wahrnehmung bzw. erwarteten und tatsächlichen Wahrnehmungen [Ha 92. S. 34ff].

Generell kann als durch Einsicht geleitet auch das sogenannte *Umwegverhalten* bestimmter Tiere eingestuft werden. Bei Laborversuchen läßt sich z.B. nachweisen, daß die von einer Ratte in einem Labyrinth gelernten Wege nicht einfach als im Gedächtnis gespeicherte starre Ketten bestimmter Reiz-Reaktionsfolgen verstanden werden können. Unter Umständen verhält sich die Ratte so, als hätte sie einen gesamten »Überblick« und »kenne« die räumliche *Struktur* des Labyrinths. Ist z.B. der direkte Weg zum Ziel ausnahmsweise versperrt, so schlägt sie fast ohne Verzögerung einen anderen, längeren Weg ein. Darauf stützt sich die Annahme, daß Ratten eine *kognitive Karte* der Umgebung entwickeln, in der sie sich aufhalten. Interessant und offen ist hier wiederum die Frage, ob es Lebewesen gibt, die einer solchen kognitiven Kartierung grundsätzlich nicht fähig sind (sind z.B. Schlangen oder Plattwürmer dazu fähig?), und ob eine notwendige (aber vielleicht nicht hinreichende) Bedingung dafür die Herausbildung von Nervenzellen ist.

## 2.5 Körperbezogene Deutung von »Konzepten«

Um mit (einigen) spezifisch menschlichen kognitiven Leistungen im Sinne unserer minimierenden Zielsetzung umzugehen, schlagen wir in groben Zügen folgendes vor. Es werden Leistungen betrachtet, die mit räumlichen Verhalten (und letztlich kognitiver Kartierung) zu tun haben. Dabei werden Konzepte, die in der Beschreibungssprache benutzt und/oder als innere Strukturen angenommen werden, zuerst soweit wie möglich vereinfacht. Bei dieser Reduktion spielen Symmetrieeigenschaften eine leitende Rolle, die an dem betreffenden Körper und der Struktur der Umgebung (einschließlich der Lage und Bewegung anderer Körper) festgemacht werden [dP 97].

Ein u.E. einleuchtendes Beispiel haben wir bereits in [dPM 94, S. 6f] formuliert, in welchem die Rückführbarkeit des relationalen Konzeptes »Zwischen« (als dreistelliges Prädikat) auf primitive (objektsprachlich als einstellige Prädikate deutbare) Verhal-

tenselemente (etwa sich links/rechts drehen, vorwärts/rückwärts gehen) gezeigt wird, welche mit Hilfe bilateraler Symmetrie bzw. mit der Asymmetrie zwischen Vorne und Hinten festgelegt werden können. In dem gleichen Zusammenhang haben wir auch ausgeführt, daß die Benutzung von Werkzeugen (als Körperverlängerungen im weitesten Sinne) und Körperteilen eine wichtige Rolle bei der Komplexität der Konzeptbildung spielen und insofern auch Anhaltspunkte für komplexitätsreduzierende Schritte liefern dürften<sup>11</sup>.

### 3. Grundlagenfragen

An dieser Stelle<sup>12</sup> möchten wir einige fundamentale Probleme im Überblick zusammenstellen und erläutern, die in verschiedenen Zusammenhängen bereits identifiziert worden sind und deren theoretische Relevanz für unsere Fragestellung unbestritten ist (bzw. sein sollte). Diese Grundlagenfragen müssen auf jeden Fall und in allen in den vorhergehenden Abschnitten angesprochenen Konkretionen berücksichtigt (und beantwortet) werden, wenn man über eine rein technische Wertschöpfung hinaus eine informatisch-wissenschaftliche anstrebt.

#### 3.1 Architektur der Komplexität

Herbert A. Simon hat in einem Aufsatz über "The Architecture of Complexity" [Si 85, S. 192-229] die These aufgestellt,

that complexity frequently takes the form of hierarchy and that hierarchic systems have some common properties independent of their specific content. Hierarchy [...] is one of the central structural schemes that the architect of complexity uses [Si 85, S. 196].

Er führt Beispiele für biologische, physikalische, soziale und symbolische Systeme an (ein Buch kann z. B. als ein hierarchisches System gesehen werden: Kapitel bestehen aus Sektionen, Sektionen aus Paragraphen usw.). Hierarchisierung ist nach Simon als *das* leitende Prinzip biologischer Evolution überhaupt anzusehen. Komplexe Systeme sind dann am stabilsten, wenn sie hierarchisch aufgebaut sind. „The time required for the evolution of a complex form from simple elements depends critically on the numbers and distribution of potential intermediate stable forms“ [Si 85, S. 202]. Die Stabilität hierarchischer Strukturen hängt wesentlich mit einer entscheidenden Eigenschaft zusammen, die der Autor mit dem Begriff der »annähernden Dekomponierbarkeit« beschreibt [Si 85, S. 209]:

In hierarchic systems we can distinguish between the interactions *among* subsystems, on the one hand, and the interactions *within* subsystems - that is, among the parts of those subsystems - on the other. The interactions at

---

<sup>11</sup> Zum Thema Konzeptbildung aus Robotersicht und den damit verbundenen Problemen vgl. auch [Mü 97].

<sup>12</sup> Dieser Abschnitt enthält ergänzte bzw. (leicht) veränderte Teile aus Kapitel 4 von [dP 93].

the different levels may be, and often will be, of different orders of magnitude.

In einer formalen Organisation wie einer Firma wird z. B. die Interaktion zwischen zwei Angestellten derselben Abteilung im Durchschnitt größer sein als die Interaktion zwischen Angestellten unterschiedlicher Abteilungen. Ähnlich sind die intermolekularen Kräfte in organischen Stoffen im allgemeinen schwächer als die molekularen Kräfte, und diese schwächer als die atomaren Kräfte. *Diejenigen Systeme, in welchen die Interaktionen zwischen den Subsystemen zwar nicht vernachlässigt werden können, aber schwach sind (weak but not negligible), bezeichnet Simon als annähernd dekomponierbar (nearly decomposable).* Solche Systeme weisen folgende Eigenschaften bzw. Vorteile auf:

- (1) the short-run behavior of each of the component subsystems is approximately independent of the short-run behavior of the other components;
- (2) in the long run the behavior of any one of the components depends in only an aggregate way on the behavior of the other components.

Als ein (sehr) einfaches Beispiel führt der Autor die Art und Weise an, wie Temperaturschwankungen innerhalb eines Hauses stattfinden können. Falls ein Haus in mehrere voneinander durch gut isolierende Wände getrennte Zimmer unterteilt ist, so gleichen sich die Temperaturen innerhalb der einzelnen Zimmer zunächst unabhängig voneinander und schneller aus als innerhalb des ganzen Hauses. Auf langer Sicht gehen die Temperaturunterschiede der einzelnen Bereiche eines und desselben Zimmers nur »aggregiert« (als Durchschnittsfaktoren) in die End-Temperatur des Hauses ein [Si 85, S. 210ff].

Da nun der Prozeß der Evolution eher komplexe Objekte hervorbringt, die annähernd dekomponierbar sind, und da kognitive Systeme auch als Produkte der Evolution angesehen werden können, scheint die Annahme berechtigt, daß sie sich in eine modulare Hierarchie zerlegen lassen.

## 3.2 Komplexität der Architektur

Daß komplexe, insbesondere kognitive Systeme in kleinere, mehr oder weniger unabhängige Untersysteme gegliedert werden können, ist in der KI und Kognitionswissenschaft eine relativ unumstrittene Annahme. So spricht z.B. J. Fodor in seinem Buch mit dem bezeichnenden Titel „The Modularity of Mind“ [Fo 83] von der *Existenz* von „informationally encapsulated“ Modulen der menschlichen Intelligenz. Die umstrittene Sache der Angelegenheit ist u.E. nicht so sehr die prinzipielle Modularisierbarkeit, sondern die *Anzahl und die Art der Organisation der Module*, mit anderen Worten, die Komplexität der Architektur. Zwei eminente Vertreter einer nicht hierarchischen Auffassung der Architektur kognitiver Systeme sind Minsky [Mi 86] und Papert [Pa 82, insb. S. 193 ff]. Sie liegen mit ihrer (bereits erwähnten) Theorie der „Gesellschaft des Geistes“ (society of mind) auf der *heterarchischen* Linie. Die Auto-

ren beziehen sich zwar unseres Wissens nicht explizit ablehnend auf Simon's Theorie der annähernden Dekomponierbarkeit, ihr Agentenmodell, das sie vorwiegend exemplarisch darstellen, ist aber nicht nur viel feinkörniger und flacher als viele andere, sondern erinnert zuweilen eher an das Chaos und die Kontroverse wettstreitender Agenten als an ein geordnetes, hierarchisch geregeltes Zusammenspiel. Der „Grundsatz der erkenntnistheoretischen Modularität“, daß Kognition (im Sinne von Wissen) in Modulen aufgeteilt werden kann [Pa 82, S. 210], gilt aber auch dort uneingeschränkt.

Dem modularen bzw. hierarchischen Dekompositionsprinzip entschieden konträre Positionen werden aber auch vertreten. So sind für D. Hofstadter kognitive Funktionen nur Epiphänomene, emergente (sogar nicht berechenbare) Eigenschaften subkognitiver Prozesse, und: „Not cognition, but subcognition, is computational“ [Ho 82]<sup>13</sup>. R. Brooks hält die Möglichkeit einer passenden *top down* funktionalen Dekomposition von kognitiven Systemen für unmöglich oder verfrüht. Unter dem Titel *Intelligence from the bottom up* plädiert er für eine »evolutionäre«, robotikorientierte KI und äußert sich folgendermassen zum Thema Modularisierbarkeit:

True intelligence requires a vast repertoire of background capabilities, experience and knowledge (however these terms may be defined). Such a system can not be designed and built as a single amorphous lump. It must have components. The same is true of existing AI systems. But true intelligence is such a complex thing that one can not expect the parts to be built separately, put together and have the whole thing work. We are in such a state of ignorance that it is unlikely we could make the right functional decomposition now. Instead we must develop a way of incrementally building intelligence ... *A key problem then is how to perform a decomposition which leads to an incremental path to intelligence* [die kursive Hervorhebung ist von uns]. [Bro 86, S. 5]

Brooks formuliert damit ein Anliegen, das unserem Ansatz ziemlich genau zu entsprechen scheint. Seine weiteren Arbeiten zur sogenannten *subsumption architecture* ([Bro 86a], [Bro 91]), die einen inkrementellen Roboterbau ermöglichen soll, und die entwicklungstechnischen Umsetzungen zeigen allerdings, daß er keine (in unserem Sinne diachronisch orientierte) Minimierung praktiziert bzw. für erforderlich hält<sup>14</sup>. Damit sind aber die Chancen für eine auch nur schattenhafte Erfüllung des bereits erwähnten Prinzips des „evolutionären Schneeballeffekts“ (vgl. § 1.4) mehr als fraglich.

### 3.3 Probleme der Modularisierung

Ist das Baukastenprinzip als (Re)Konstruktionsmethode für kognitive Systeme nicht grundsätzlich unzulänglich? Ist ein Aufbau von innen (unten) nach außen

---

<sup>13</sup> Diese Position ist, wie der Autor selbst in einem Postscriptum bemerkt, weitgehend im Einklang mit den Grundannahmen, die im Bereich *Konnektionismus* gemacht werden. Dementsprechend wird sie gewürdigt, z .B. von Clark [Cl 97].

<sup>14</sup> Er träumt ja inzwischen auch von der Realisierung menschenähnlicher Roboter [BroSt 93].

(oben) geeigneter als ein Aufbau von außen (oben) nach innen (unten)? Oder ist es etwa besser, in der Mitte anzufangen, und von dort gleichzeitig in alle anderen Richtungen vorzugehen? Vielleicht hängen Innen und Außen gar nicht *linear*, sondern ganz anders, etwa *kreis-* oder *spiralförmig* oder sogar *möbiusbandartig* (in diesem Fall wäre die Unterscheidung nur eine Frage der lokalen Perspektive), zusammen. Im folgenden versuchen wir systematisch die problematischen Aspekte zu besprechen, die sich mit der Annahme der Modularisierbarkeit von kognitiven Systemen ergeben.

### 3.3.1 Entwicklung von Komplexität

In dem weiter oben zitierten Ausschnitt spricht Brooks ein fundamentales Problem an: es reicht nicht, kognitive Systeme als komplexe Module zu beschreiben, man muß auch sagen, wie die Komplexität zustandekommt. Die Vorstellung, daß man ein kognitives System ein für allemal wie ein Uhrwerk entwerfen und zusammensetzen kann und dann nur aufzuziehen braucht, damit es läuft, ist nicht adäquat. Man kann zumindest nicht erwarten, daß ein solches System etwas wesentlich anderes tut, als rein repetitive Aufgaben (wie die Zeit zu messen). Im gleichen Sinne spricht Papert von *Fragen der Entwicklung*, welche ein *dynamisches* Modell für die mögliche *Entstehung und Veränderungen* von kognitiven *Strukturen* selbst verlangen [Pa 82, S. 205]. Man beachte nun, daß in dieser Hinsicht bis heute noch keine überzeugenden Ergebnisse vorliegen. Selbst die auf eine „klärende“ Festlegung der Strukturen zielende Annahme, daß ein kognitives System »repräsentationistisch«, d. h. auf der Grundlage sich veränderbarer interner Repräsentationen von sich selbst und seiner Umwelt arbeitet, ist noch umstritten<sup>15</sup>.

Die Entwicklung von Komplexität ist allerdings nicht nur ein kognitionswissenschaftliches Problem, sondern ein allgemeines Problem. Wie die fundamentalen Begriffe »Ordnung«, »Organisation« und »Komplexität« einerseits genau voneinander abgegrenzt werden können und darauf basierend, andererseits z.B. Prozesse der »Selbstorganisation« bzw. der »Emergenz« neuer, globaler Eigenschaften aus lokalen Interaktionen beschrieben werden können, ist trotz ansehnlicher Literatur weit entfernt von einer einheitlichen Betrachtung<sup>16</sup>. Es ist z. B. unklar, ob (biologische) Organisati-

---

<sup>15</sup> Vgl. z. B. [MaVa 87, S. 145 ff], [Var 88, S. 88 ff] oder Peschl [Pe 90, S. 30 f], der sogar von 'repräsentationsfreier Repräsentation' redet (und damit leider einen Beitrag zu der inzwischen fast inflatorischen Entwertung bestimmter KI-Begriffe (hier Repräsentation) leistet).

<sup>16</sup> Zur Verdeutlichung (eines Aspektes) der Problematik möge folgendes Zitat aus [KüKr 97a, S. 165] dienen. Es geht um »Selbstorganisation«: „Rein formal sind alle Systeme selbstorganisierend, bei denen die Ursachen für Systemveränderungen im System selbst liegen und externe Einflüsse auf die Systemdynamik nicht entscheidend sind. Mit anderen Worten, autonome Systeme sind, sofern sie überhaupt organisiert sind, selbstorganisiert. Diese Definition [...] läßt zu, daß jedes fremdorganisierte System durch eine Verschiebung der Grenzen, bei der die außerhalb liegenden Quellen der Veränderung in das System integriert werden, zu einem selbstorganisierendem System wird.“ Das ist natürlich keine »fruchtbare« Definition, weil die Verschiebung des »Randes« der Willkür des Beobachters unterworfen ist. Wenn man allerdings verlangt, daß „das System sich selbst von seiner Umwelt abgrenzt, das heißt einen Rand bildet [...]

on und Komplexität überhaupt *meßbar*, z. B. *additive* Größen sind: Sind etwa zwei Bakterien doppelt so komplex (oder doppelt so organisiert) wie ein Bakterium? Ist die Interaktion einer Katze mit einer Maus komplexer als die Interaktion einer Katze mit einem Hund? (Für eine Vertiefung der relativen Rollen von Komplexität, Chaos, Nicht-Linearität, Selbstorganisation aus verschiedenen Sichten vgl. die Sammelbände von [KüKr 97] und [Kü 96] sowie [Dav 87], insb. Kap. 3-6).

In demselben Sinne stellt sich die Frage, ob Kognition entwicklungsmäßig etwa ein kumulatives »Gesamt«-Phänomen ist, vergleichbar beispielsweise mit dem Zustandekommen des Gewichtes eines lebenden Systems. Oder ist sie etwas, das quantitativ nicht reduzierbar, sondern nur qualitativ als ganzheitliche Eigenschaft verstehbar ist? Man denke hier etwa an die charakteristische Eigenschaft von Rasterbildern. Ein Rasterbild, z. B. die Fernsehaufnahme eines menschlichen Gesichtes, läßt sich nicht aus der lokalen Untersuchung einzelner Punkte ableiten. Erst wenn man aus einer angemessenen Entfernung die Punkte als Gesamtheit betrachtet, wird das Bild sichtbar. Das Gesicht scheint also keine Eigenschaft der elementaren Bestandteile des Bildes, d. h. der jeweiligen Rasterpunkte zu sein, sondern ihrer (globalen) Anordnung<sup>17</sup>. Muß man also in Analogie dazu eher nach *Kognitionsmustern* als nach *Kognitionsbausteinen* suchen?

### 3.3.2 Introspektion bzw. Reflexivität

Ein weiteres grundlegendes, von einer Lösung weit entferntes Problem im Zusammenhang mit der Modularitätsannahme ist relativ früh von Minsky formuliert worden. Er spricht von Computerprogrammen, die a) als *Modelle* von *Minds* betrachtet werden können, und b) als *Teil* auch ein Modell von sich selbst enthalten können:

We use the term 'model' in the following sense: To an observer B, an object A\* is a model of an object A to the extent that B can use A\* to answer questions that interest him about A. [Mi 68, S. 426]

The notion of "part" is more complicated for things like computer programs than for ordinary physical objects. A single conditional branch makes it possible for a program to behave, functionally, like two very different machines in different circumstances, yet using almost (or exactly) the same sets of instructions. The notion of a machine containing a model of itself is also complicated. [... Such a machine] can use the internal description to calculate what it itself would do under some hypothetical circumstance. [...] With the possibility of this sort of *introspection*, [die kursive Hervor-

---

und ihn aufrechterhält“ (ibidem, S. 166), so ist man wieder (im Sinne einer Umdefinition) am Anfang des Problems.

<sup>17</sup> Ein interessantes (extremes) Beispiel ist das Block-, d. h. grob gerastertes Porträt von A. Lincoln (vgl. [Fr 89, S. 80 und 133] und [StKr 97, S. 142f]), das zeigt, daß die Anzahl der Raster (die Auflösung der Bilder) für den Gesamteindruck nur eine relative Rolle spielt. In diese Fragestellung spielt die Gestaltproblematik mit ihrem ganzheitlichen Ansatz sicherlich eine wesentliche Rolle (vgl. z.B [Me 75]).

hebung ist von uns] the boundaries between parts, things and models become very hard to understand. [Mi 68, S. 430]

Obwohl die Forschung in diesem Bereich, den man mit dem Stichwort *Reflexion* bzw. *Reflexivität* umschreiben kann, in letzter Zeit intensiver betrieben wird [MaeNa 87], ist hier immer noch viel Klärungsarbeit notwendig.

Das Reflexionsproblem ist entscheidend im Zusammenhang mit der Architekturproblematik. Die meisten KI-Systeme verfügen nicht über ein Modell von sich selbst, kennen z.B. nicht ihre eigenen Grenzen: Ich (als Mensch) weiß, daß ich nicht (wie ein Vogel) fliegen kann. »Weiß« das eine Katze oder gar eine Ameise, die (vermutlich überhaupt) keine Vögel wahrnehmen kann? Entscheidend ist auch die Frage, ob die »formale« Leistungskraft (Mächtigkeit im verarbeitungstechnischen Sinne) von Reflexion mit der Mächtigkeit des (programmiersprachlichen) Rekursionsprinzips vergleichbar ist. Wäre das der Fall, so könnte man behaupten, daß ein nicht reflexives (introspektives) System bestimmte (kognitive) Leistungen bzw. Funktionen überhaupt nicht erbringen kann. Varela sieht z.B. Reflexivität, die er kurz als Erfahrung des Selbstbezuges („l'expérience de se référer à soi“) bzw. Selbstbeschreibungsfähigkeit („capacité d'autodescription“) bezeichnet, in einer notwendigen Verbindung zur (menschlichen) Sprachfähigkeit: „La réflexivité est quelque chose d'absolument crucial, c'est la grande mutation qui se produit avec l'apparition du langage chez l'homme“ [Ke 98].

Die Trag- und Reichweite dieser Frage wird klar, wenn einige Wissenschaftler im Recht sind, die Rekursion als biologisch verankertes Prinzip betrachten. „Because DNA chemistry is like computing, it also makes use of recursion. [...] the DNA blueprint not only has to specify designs for the tools, but also for the tools that make the tools. And it must provide instructions for how to read the instructions.“ [CoSt 95, S. 80f]. Diese Autoren sehen in der Rekursion über die genetischen Aspekte hinaus ein allgemeines, auch auf der Ebene gesellschaftlicher Organisationen, z.B. auf Produktionsebenen geltendes Prinzip: „Modern factory technology is also recursive: The machine tools used to make cars are themselves made by machine tools“ [ib., S. 82]. Aus unserer Sicht stellt sich die Frage, wie der Übergang von der tief biologisch gelagerten Rekursion in höhere, (möglicherweise) auch rekursiv funktionierende kognitive Schichten stattfindet.

### **3.3.3 Zirkularität**

Ein weiteres fundamentales Problem, das der Modularisierbarkeit von kognitiven Systemen anhaftet, ist die sich aus der Beschaffenheit der Module ergebende potentielle Zirkularität. Dies ist besonders deutlich, wenn man ein kognitives System als »Agent« versteht: Solange man die Bausteine (Module) des Systems auch als Agenten darstellt, dreht sich die Theorie im Kreis. Sie erklärt das Verhalten von kognitiven Systemen mit dem Verhalten von kognitiven Systemen [Pa 82, S. 208]. Sehr treffend



spricht J. Haugeland von *Homunkulus-Schachtelung* auf verschiedenen Ebenen [Hau 87, S. 98]. Der Aufbau von außen nach innen (die Schachtelung) funktioniert allerdings nur, wenn die inneren Homunkuli immer dümmer werden (bis hinunter auf Null); und umgekehrt funktioniert der Aufbau von innen nach außen nur dann, wenn man, bei Null startend, darüber hinausgehen und fortschreiten kann. Insgesamt erinnert das in etwa an das geometrische Paradoxon der Linie, die als Objekt mit einer Extension aus lauter Objekten (Punkten) »zusammengesetzt« ist, die selbst gar keine Extension haben. Was in der Mathematik zulässig und möglich ist, muß keine biologische Grundlage haben.

Als interessant zu werten ist in dieser Hinsicht der Versuch von Konstruktivisten, die Zirkularität letztlich gar nicht als Problem, sondern als Vorteil ansehen. Varela spricht z.B. provokativ von *circuli virtuosi* bzw. *fructuosi* anstatt von *circuli vitiosi*, allerdings nur in einem erheblich erweiterten, (neuro)biologisch und physikalisch gemeinten Kontext ([Va 88], [Va 90]). Aus molekularbiologischer Sicht stellt die Hyperzyklen-Theorie von M. Eigen (vgl. [Ku 92, S. 34ff]) Zirkularität auch in ein anderes (positives) Licht. Vielleicht sind damit die Stellen identifiziert, an welchen Zirkularität und Rekursion *anfangen*.

### 3.3.4 Skalierbarkeit

Eng verbunden mit dem Problem der Zirkularität und Reflexion ist ein anderer Aspekt, den wir Skalierbarkeit von kognitiven Systemen nennen möchten. Darunter ist grob zu verstehen, ob für Kognition, sofern für sie überhaupt ein »Maßstab« gefunden werden kann, eine Varianz oder Invarianz desselben gegeben ist: Sind kognitive Bausteine und Funktionen (oder Muster) im Kleinen (»Mikro«-Kognition) vergleichbar mit Bausteinen und Funktionen (oder Muster) im Großen (»Makro«-Kognition)? Können größere Teile »gesetzmäßig« aus kleineren zusammengesetzt werden? Diese Fragen sind naturwissenschaftlich relevant. In der Physik weiß man z.B., daß viele Gesetze Symmetrieeigenschaften besitzen, d. h. invariant bzgl. einiger Transformationen (Translationen, Drehungen, zeitlicher Verschiebungen usw.) sind. Bzgl. einer Änderung des *räumlichen* Maßstabs (sogenannter Ähnlichkeitstransformationen) sind physikalische Gesetze allerdings *nicht* invariant. Das heißt, daß das rein mathematisch-geometrische Ähnlichkeitsprinzip nicht auf die physikalische Gesetze in Strenge anwendbar ist [Ta 93, S. 105ff]. Dies läßt (in Analogie) in unserem Kontext die Frage als berechtigt erscheinen, ob z.B. für die (möglicherweise) abstufbaren Bereiche der Kognition (von Mikro bis Makro) jeweils andere Aufbau-Prinzipien gelten. Ein computer-technisch gesehen einfacher, durchgehender rekursiver Aufbau könnte prinzipiell inadäquat sein. Was symmetrisch auf der (äußeren) Ebene einer Verhaltens- oder Formbetrachtung erscheint ([dP 97], [PM 97]), könnte z.B. keine auf Symmetrien fußende Entsprechung in tieferen Systemschichten haben.

Aus mathematischer Sicht ist die Frage weiter präzisierbar. Selbstähnlichkeit ist bekanntlich fundamental für die Geometrie. Grundbausteine (z. B. Strecken und Quadrate) können nach Potenzgesetzen zusammengefügt werden, um das »Maß« größerer Objekte zu ergeben. Nach Entdeckung der fraktalen Geometrie [Ma 77] läßt sich nun folgendes unterscheiden. Für die Beschreibung von Artefakten (wie z.B. architektonische Werke) reicht in der Regel ein Rückgriff auf die Euklidische Geometrie aus, wobei sowohl die Grundbausteine als auch die Regeln für ihre Komposition »einfach« sind. Versucht man allerdings auf die gleiche Art und Weise, d. h. euklidisch, »natürliche« Formen (wie z.B. ein Baum oder eine Wolke) zu beschreiben, so erweist sich der mathematische Apparat als sehr schwerfällig. „Eine Euklidische Beschreibung des schwarzen Milzfarns könnte z.B. ein Polygon mit Tausenden angepaßter Parameter erfordern“ [HaSu 96, S. 3]. Eine fraktale, also nicht Euklidische, Beschreibung natürlicher Formen vereinfacht das Problem in dem Sinne, daß die Komplexität von den Kompositionsregeln in die Grundbausteine selbst verlagert wird. Das Erzeugungsverfahren ist dabei (ganz einfach) Rekursion! „Die Zusammensetzung von Fraktalen aus ihren natürlichen Bausteinen, nämlich kleineren selbstähnlichen Fraktalen, scheint wirklich einen wesentlichen Bestandteil der Wachstums- und Verzweigungsgesetze der Natur selbst zu erfassen“ [HaSu 96, S. 3]. In Analogie dazu kann nun die Frage über die *Dimension* von Kognition (qua natürlichem Phänomen) gestellt werden, und ob diese fraktaler (komplizierte Leistungen sind mit einfachen Verfahren aus komplizierten Bausteinen zusammengesetzt) oder Euklidischer Natur ist (komplexe Leistungen entstehen durch komplizierte Verfahren, die über einfache Bausteine arbeiten). Die Annahme (und Hoffnung), daß sowohl Bausteine als auch Verfahren einfach sind, könnte trügerisch sein.

### 3.4 Begründungsproblematik

Auf einer anderen aber ebenso grundsätzlichen und allgemeinen Ebene, die komplementär zu derjenigen ist, die wir zuletzt betrachtet haben, wird mit dem Zirkularitäts- bzw. Skalierungsproblem die Frage der Erklärung von Funktionen und Strukturen im Sinne einer *Begründung* angesprochen. Die auf einer sehr alten philosophischen Tradition beruhende Begründungsproblematik ist hinsichtlich ihrer methodischen Schwierigkeiten von H. Albert auf die Form des sogenannten *Münchhausen-Trilemmas* (in Anspielung an das in informatischen Kreisen etwas bekanntere bootstrapping-Dilemma des berühmten Lügenbarons) gebracht worden [Al 80, S. 8ff]. Etwas verallgemeinert (Albert hat eher »logisch deduktive« Begründungsbemühungen im Auge gehabt) besagt das Trilemma folgendes: Wenn man versucht (die Geltung von) etwas auf etwas anderes zurückzuführen, so steht man vor drei Alternativen, die sämtlich (aber nicht gleichermaßen) angreifbar sind:

- a) einem *infiniten Regreß*, in dessen Rahmen die Kette der zur Begründung anstehenden »Dinge« nicht abbricht,

- b) einem *Zirkel*, bei dem der Regreß dadurch scheinbar vermieden wird, daß ein und dasselbe »Ding« als seine eigene Begründungsbasis erscheint, d. h. an zwei Stellen der Begründungskette auftritt,
- c) der dogmatischen *Festsetzung einer Begründungsbasis*, die ex definitione einer Begründung nicht bedürfen soll.

Wir bewegen uns mit unseren Überlegungen zur biologischen Fundierung von Kognition innerhalb dieser letzten Schwierigkeit.

S. Papert ist sich z.B. dieser Problematik völlig bewußt: er sieht für seine Theorie des Geistes als Gesellschaft von Agenten (*society of mind*) eine ausreichende, »sparsame« und *nicht zirkuläre* Erklärung bereits in einem »speziellen computertechnischen Grundsatz« verkörpert, nämlich der Theorie der reinen Prozeduren, d. h. Prozeduren, die abtrennbare *elementare* Funktionen realisieren und wie Moduln für den Bau komplexer, »debuggbarer« Einheiten benutzt werden können [Pa 82, S. 209].

### 3.4.1 Beispiel: Zielorientiertheit und Rückkopplung

Im Unterschied bzw. in Ergänzung zur eben erwähnten Papertschen Position (wobei wir die Frage der Abtrenn- und Zusammensetzbarkeit elementarer Einheiten, die biologisch basiert sind, als offen betrachten) halten wir eine Erklärung nur dann im Sinne einer Begründung für zumindest ansatzweise akzeptabel, wenn in ihr ein Übergang von der *matter* zum *mind* unmittelbar möglich ist. Diesen Standpunkt können wir zur Zeit nicht allgemein, sondern nur exemplarisch charakterisieren. Das kybernetische Prinzip der *negativen Rückkopplung* (s. z. B. [Cu 75, S. 126ff]) kann z. B. als *Erklärung* für zweckgerichtetes Handeln fungieren. Wenn ein Agent ein bestimmtes Ergebnis erzielen will und durch irgendeinen störenden Einfluß daran gehindert wird, versucht er, diesem Einfluß entgegenzuwirken. Wenn er z. B. eine gleichbleibende Temperatur in seinem Zimmer haben möchte, wird er dazu neigen, die Heizung niedriger bzw. höherzudrehen, je nachdem, ob es im Raum wärmer oder kühler wird. Genauso *verhält sich* ein Thermostat [Hau 87, S. 149]. In einem Thermostat als physikalischem Gerät ist also ein Prinzip *verkörpert*, das auch als (notwendiges, wenn auch nicht hinreichendes) Bauprinzip eines kognitiven Systems denkbar ist. *Negative Rückkopplung* kann insofern nach unserem Verständnis als (naturwissenschaftlich) angemessene *Begründung* für eine bestimmte kognitive Funktion (die Zielorientiertheit) betrachtet werden<sup>18</sup>. Ähnlich argumentiert N. Bischof [Bi 89, S. 101ff], der es als ein Verdienst der Biokybernetik ansieht, die biologisch vernünftig interpretierbare Annahme von Finalursachen von der vitalistischen Unterstellung, sie seien immate-

---

<sup>18</sup> Ein expliziter Zusammenhang zwischen Zielorientiertheit und negativer Rückkopplung wurde in der Kybernetik bereits sehr früh festgestellt: „All purposeful behavior may be considered to require negative feed-back“ [RWB 43, S. 19]. Die Autoren führten interessante Differenzierungen in Abhängigkeit der Stationarität des Ziels und der „predictive“ Reichweite des Verhaltens (Unterschied z.B. zwischen einfachen Tropismen und Verfolgung bewegter Ziele mit der Möglichkeit bzw. Notwendigkeit ihre zukünftigen Positionen zu extrapolieren).

rieller Natur, befreit zu haben. Der aristotelische Begriff der Zielstrebigkeit kann als Eigenschaft von Regelkreisstrukturen entmystifiziert werden: Zielstrebigkeit bedeutet *Homöostase*, d. h. Stabilisierung von Systemzuständen um Sollwerte<sup>19</sup>. Das Prinzip der Zweck-Mittel-Analyse [NeSi 76], das in der KI als allgemeines Modell zweckgerichteten Vorgehens gilt (Erreichung eines Zieles durch Minimierung der Differenzen zwischen Anfangs- und Zielsituation) bietet keine so direkt »anfaßbare« Grundlage.

### 3.4.1 Beispiel Sprache

Das Begründungsproblem betrifft auf einer sehr fundamentalen Art und Weise Ansätze, die sprachorientiert vorgehen, darunter insbesondere logisch orientierte Erklärungsmodelle, die in dem Sinne »sprachimmanent« sind, daß sie Sprache sozusagen gleichermaßen als Ausgang, Mittel und Zielpunkt haben. Unsere Sicht ist, daß Sprache bzw. Logik an sich sicherlich als höchste und erklärungsbedürftigste menschliche kognitive Leistungen anzusehen sind, daß sie aber nicht im Sinne einer (minimalistischen) Begründung als Basis für Anderes insbesondere nicht für sich selbst geeignet sind. Die dynamische, zeitbehafte Entwicklung von (nicht nur kognitiven) Strukturen ist logisch-formal nicht »erklärbar«. Oder ist etwa eine endliche Kette von Aussagen denkbar, die einen Menschenaffen mit Darwin verbindet und nach den formalen Regeln aufgebaut ist: der Sohn eines Affens ist ein Affe, der Vater eines Menschen ist ein Mensch?<sup>20</sup>

Evident ist die grundlegende Begründungs- und Erklärungsunfähigkeit der Logik wegen des Regresses, der schon im Zuge der Festlegung der Bedeutung einfacher Sprachmittel auftritt, die für logische Beschreibungen erforderlich sind. Die Gültigkeit von Aussagen oder Schlüssen (*Theoremen*) auf der Ebene einer bestimmten Objektsprache ergibt sich nur dadurch, daß genau entsprechende Schlüsse in einer entsprechenden mächtigeren Metasprache als gültig vollzogen werden [Tar 77, S. 244ff]<sup>21</sup>. Wollte man diese auch »objektivieren«, so brauchte man eine Meta-Meta-

---

<sup>19</sup> Etwas unklar ist (für uns) in diesem Zusammenhang der Stellenwert *positiver* Rückkopplung, die eintritt, wenn die Veränderung der zu regelnden Größe nicht entgegengesetzt, sondern in der gleichen Richtung der Veränderung der Störgröße läuft. Sie ist vermutlich genauso notwendig wie die negative Rückkopplung, die für Stabilisierung sorgt, weil sie Systemübergänge erlaubt. Wie sollte sonst, d. h. ohne positive Rückkopplung, aus einer Kaulquappe je ein Frosch werden können? Man beachte ferner, daß Rückkopplung aus kybernetischer Sicht als *einfachstes* materielles Korrelat von Verhaltensanpassung gesehen wird. Homöostasis wird darüber hinaus auf den Begriff der Stufenfunktionen gestützt, die mit einem *komplizierteren* Prinzip verknüpft sind, nämlich dem des Wahlvermögens oder der Spontaneität. Ein Homöostat ist (im Unterschied zu einem Thermostaten) *ultrastabil*, d. h. in der Lage, sein »inneres Milieu«, also einen bestimmten Zustand, trotz vielfältiger innerer und äußerer Störungen aufrechtzuerhalten [Cu 75, S. 192 f].

<sup>20</sup> Dieses Beispiel ist aus [St 87, S. 73] entnommen.

<sup>21</sup> An einem Beispiel dargestellt: *Sind A und B Aussagen, so heißt "A v B" genau dann wahr, wenn A wahr ist oder wenn B wahr ist.* Hier wird die Semantik des Junktors "v" (oder) festgeschrieben, indem man die natürliche Sprache Deutsch als Metasprache benutzt (darauf weist die kursive Hervorhebung im Text), deren Semantik schon als bekannt vorausgesetzt wird.

sprache, und so weiter, ohne daß ein Ende für eine so aufzubauende Hierarchie von Sprachen abzusehen wäre. Dies macht den Charakter der Nichtabgeschlossenheit von Sprache überhaupt deutlich und legt eine negative Antwort auf die grundsätzliche Frage nahe, ob Sprache mit Sprache überhaupt *begründet* werden kann. Es ist wohl *intuitiv* klar, daß, genauso wie die letzte Begründung eines Satzes nicht ohne unendlichen Regress ein Satz sein kann, eine Sprache nicht durch eine andere Sprache begründet werden kann. Was die moderne Logik tut, ist nichts anderes als eine Form unvollständiger *Übersetzungssemantik*. Dadurch erklärt sie nichts und setzt im Gegenteil alles voraus. Gerade die Unzufriedenheit mit dieser Situation hat in letzter Zeit zu einer neuen, vielversprechenderen Orientierung der Logik-Forschung geführt, in der man nach einer alternativen, d. h. *sprachunabhängigen* Fundierung der Logik im (*dialogischen*) *Handeln* gesucht hat [Lo 74, S. 60ff].

Daß speziell logisch-mathematische Strukturen nicht ausschließlich von sprachlichen Formen abgeleitet werden können, wird aus der Sicht psychologischer Grundlagenforschung auch von J. Piaget behauptet. Der Autor weist auf den *circulus vitiosus* hin, der der Abbildtheorie der Erkenntnis zugrundeliegt, die Erkenntnis als ein passiv empfangendes Abbild der Realität auffaßt, und vertritt die Meinung, daß Erkennen eher mit *Einwirken* auf Objekte zu tun hat. Nach seiner Hypothese „ist das, wovon abstrahiert wird, nicht das Objekt, auf das eingewirkt wird, sondern das einwirkende *Handeln* [die kursive Hervorhebung ist von uns]“ [Pi 73, S. 24]. Diese Auffassung entspricht in der Intention dem bereits erwähnten Versuch von P. Lorenzen [Lo 74], Logik nicht auf Sprache, sondern auf dialogisches Handeln zurückzuführen. Unser Vorschlag (vgl. § 2.5), höhere Konzepte bzw. kognitive Leistungen körper- und verhaltensbezogen zu interpretieren und zu reduzieren, ist auch in diesem Kontext zu sehen.

Der Aufbau eines kognitiven Systems, so kann man abschließend die obigen Überlegungen zusammenfassen, ist nicht nur eine Frage von logischen Beschreibungen und »Repräsentationen« von Strukturen und Funktionen. Er muß auf etwas anderem als nur auf Sprache beruhen.

## **Danksagung**

Für die Durchsicht früherer Versionen dieses Papiers und verschiedene Anmerkungen bin ich insbesondere B. S. Müller zu Dank verpflichtet. Für manche informelle und anregende Gespräche über einige der dargestellten Ideen bin ich auch den weiteren Kollegen aus dem Bereich »Kognitive Robotik« sowie Thomas Christaller sehr dankbar, der die Rahmenbedingungen für die Durchführung dieser Arbeit geschaffen hat.

# Literatur

- [AgBe 97] Agah, A./Bekey, G. A., Phylogenetic and Ontogenetic Learning in a Colony of Interacting Robots, in *Autonomous Robots* 4, 85-100 (1997)
- [Al 80] Albert, H., *Traktat über kritische Vernunft*, J.C.B. Mohr, Tübingen, 1980
- [ApDö 80] Apfelbach, R./Döhl, J., *Verhaltensforschung*, 3. Auflage, UTB 210, Stuttgart, 1980
- [ArEyMe 97] Arnold, W./Eysenck, H. J./Meili, R., *Lexikon der Psychologie*, Bechtermünz Verlag, Augsburg, 1997
- [BeDiV 95] Bennett, C. H./DiVincenzo, D. P., Quantum Computing – Towards an engineering era?, in: *News and Views (Nature)*, Vol 377, 5 October 1995, S. 389-390
- [BG 87] Balkemore, C./Greenfield, S. (eds), *Mindwaves – Thoughts on Intelligence, Identity and Consciousness*, Basil Blackwell, Oxford, 1987
- [Bi 89] Bischof, N., Ordnung und Organisation als heuristische Prinzipien des reduktiven Denkens, in: Meier, H. (Hrsg.), *Die Herausforderung der Evolutionsbiologie*, Piper, München, 1989, 79-127
- [Bi 91] Bischof-Köhler, D., Jenseits des Rubikon – Die Entstehung spezifisch menschlicher Erkenntnisformen und ihre Auswirkung auf das Sozialverhalten, in : Fischer, E. P. (Hrsg.), *Mannheimer Forum 90/91*, Piper, München, 1991, S. 143-194
- [Bro 86] Brooks, R. A., *Achieving Artificial Intelligence Through Building Robots*, MIT AI-Memo 899, May, 1986
- [Bro 86a] Brooks, R. A., A Robust Layered Control System for a Mobile Robot, in: *IEEE Journal of Robotics and Automation*, April 1986, S. 14-23
- [Bro 91] Brooks, R. A., Challenges for Complete Creature Architectures, in: Meyer, J./Wilson, S. (eds), *From Animals to Animats*, The MIT Press, 1991, S. 434-443
- [BroSt 93] Brooks, R. A./Stein, L. A., *Building Brains for Bodies*, A.I. Memo No. 1439, MIT, 1993
- [BrSc 93] Braitenberg, V. / Schüz, A., *Allgemeine Neuroanatomie*, in: Schmidt, R. F. (Hrsg.), *Neuro- und Sinnesphysiologie*, Springer-Verlag, Berlin, 1993, S. 1-30
- [BCS 91] Beer, R. D./Chiel, H. J./Sterling, L. S., A Biological Perspective on Autonomous Agent Design, in: [Mae 91], S. 169-185
- [CaKa 95] Casti, J. L./Karlqvist, A., *Cooperation and Conflict in General Evolutionary Processes*, John Wiley & Sons, New York, 1995
- [Ch 98] Christaller, T., Cognitive Robotics: A New Approach to Artificial Intelligence in: Sugisaka, M. (Ed), *Proceedings of Third Int. Symp. on Artificial Life, and Robotics (AROB III' 98)*, Oita, Japan, 19-21, January, 1998, S. 495-498
- [Cl 97] Clark, A., *Microcognition – Philosophy, Cognitive Science, and Parallel Distributed Processing*, The MIT Press, 1993

- [Cl 97a] Clark, A., The Dynamical Challenge, in: Cognitive Science, 4 (1997) 461-481
- [Co 90] Connell, J. H., Minimalist Mobile Robotics – A Colony-style Architecture for an Artificial Creature, Academic Press, Boston, 1990
- [CoSt 95] Cohen, J./Stewart, I., The collapse of chaos, Penguin Books USA, 1995  
Deutsche Version: Chaos und Anti-Chaos – Ein Ausblick auf die Wissenschaft des 21. Jahrhunderts, dtv, München, 1997
- [Cu 75] Cube, F. von, Was ist Kybernetik, dtv Wissenschaftliche Reihe, München 1975
- [Da 95] Dawkins, R., River Out of Eden, BasicBooks, New York, 1995
- [Dav 87] Davis, P., Prinzip Chaos, Bertelsmann, München, 1987
- [DauCh 95] Dautenhahn, K./Christaller, T., Remembering, rehearsal and empathy: Towards a social and embodied cognitive psychology for artifacts, Arbeitspapiere der GMD, Nr. 956, November 1995
- [DMS 96] Dudel, J./Menzel, R./Schmidt, R. F. (Hrsg.), Neurowissenschaft: vom Molekül zur Kognition, Berlin, Springer, 1996
- [dP 93] di Primio, F., *Hybride Wissensverarbeitung*, DUV, Vieweg, 1993
- [dP 97] di Primio, F., Symmetry and Symmetry-Breaking in Multi-Agent Behavior in: Fujimura, S./Sugisaka, M. (Eds), *Proceedings of International Symposium on Artificial Life and Robotics (AROB '97)*, ISBN4-9900462-7-7, Oita University, Japan, 1997, S. 28-33
- [dPM 94] di Primio, F./Müller, B.S., Research Prospects on Cognition and Behavior, Arbeitspapiere der GMD, Nr. 831, March 1994
- [Er 98] Ernst, A. M. (Hrsg.), Digest: Kooperation und Konkurrenz, Spektrum der Wissenschaft, Heidelberg, 1/1998
- [Fo 83] Fodor, J. A., The Modularity of Mind, The MIT Press, MA, 1983
- [Fr 89] Frisby, J. P., Optische Täuschungen – Sehen • Wahrnehmen • Gedächtnis, Weltbild Verlag, Augsburg, 1989
- [FuUe 94] Fukuda, T./Ueyama, T., Cellular Robotics and Micro Robotic Systems, World Scientific, London, 1994
- [Ga 89] Gazzaniga, M. S., Das erkennende Gehirn – Entdeckungen in den Netzwerken des Geistes, Junfermann-Verlag, Paderborn, 1989
- [Gla 97] von Glaserfeld, E., Radikaler Konstruktivismus, stw 1326, Frankfurt 1997
- [Ha 92] Hassenstein, B., Klugheit – Zur Natur unserer geistigen Fähigkeiten, Piper, München, 1992
- [HaBu 90] What connectionist models learn: Learning and representation in connectionist networks, in: Behavioral and brain sciences (1990) 13, 471-518
- [Har 90] Harnad, S., The Symbol Grounding Problem, in: Physica, D 42, 1990, S. 335-346
- [HaSu 96] Hastings, H. M./Sugihara, G., Fraktale, Spektrum Akademischer Verlag, Berlin, 1996

- [Hau 87] Haugenland, J., Künstliche Intelligenz – Programmierte Vernunft?, McGraw-Hill, 1987
- [Hei 97] Heiden, U. an der, Selbstorganisation in dynamischen Systemen, in: [KüKr 97], S. 57-88
- [Ho 82] Hofstadter, D. R., Waking up from the Boolean Dream, or, Subcognition as Computation, July 1982, in: Hofstadter, D. R., Metamagical Themes: Questing for the essence of mind and Pattern, Penguin Books, 1986, 631-665
- [Jae 96] Jaeger, H., Dynamische Systeme in der Kognitionswissenschaft, in: Kognitionswissenschaft (1996) 5: 151-174
- [JL 88] Johnson-Laird, P. N., The computer and the mind – An introduction to cognitive science, Fontana Press, London, 1988
- [Jo 66] Jolly, A., Lemur Social Behavior and Primate Intelligence, in: Science, Vol. 153, 1966, S. 501-506
- [KaLi 96] Kaelbling, L. P./Littmann, M. L., Reinforcement Learning: A Survey, in: Artificial Intelligence 4 (1996) 237-285
- [Ke 98] Kempf, H., Francisco Varela: "Le cerveau n'est pas un ordinateur – C'est de l'activité permanente du corps qu'émerge le sens de son monde", in: La recherche, 308, Avril 1998
- [Kl 88] Kline, P., Psychology Exposed – Or The Emperor's New Clothes, Routledge, London, 1988
- [Kli 93] Klix, F., Erwachendes Denken – Geistige Leistungen aus evolutionspsychologischer Sicht, Spektrum Akademischer Verlag, Berlin, 1993
- [Ku 92] Kull, U., Evolution, Metzler, Hannover, 1992
- [Kü 96] Küppers, G. (Hrsg.), Chaos und Ordnung – Formen der Selbstorganisation in Natur und Gesellschaft, Reclam, Stuttgart, 1996
- [KüKr 97] Küppers, G./Krohn, W. (Hrsg.), Emergenz: Die Entstehung von Ordnung, Organisation und Bedeutung, Suhrkamp, Frankfurt, 1997
- [KüKr 97a] Küppers, G./Krohn, Zur Emergenz systemspezifischer Leistungen, in: [KüKr 97], S. 161-188
- [Lo 74] Lorenzen, P., Methodisches Denken, Suhrkamp, Frankfurt, 1974
- [Ma 77] Mandelbrot, B., Die Fraktale Geometrie der Natur, Birkhäuser, Basel, 1977
- [Mac 87] Macphail, E., Intelligence: A Comparative Perspective, in: [BG 87], S. 177-193
- [Mae 91] Maes, P. (ed), Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back, The MIT Press, Cambridge (MA), 1991
- [MaeNa 87] Maes, P./Nardi, D. (Eds.), Meta-level Architectures and Reflection, North-Holland, 1987
- [MaSm 91] Malcom, C., Smithers, T., Symbol grounding via a hybrid architecture in an autonomous assembly system, in: [Mae 91], S. 123-144
- [MaVa 87] Maturana, H. R./Varela, F. J., Der Baum der Erkenntnis, Scherz, 1987



- [McC 88] McCarthy, J., Mathematical Logic in Artificial Intelligence, in: Daedalus, Winter 1988, Artificial Intelligence, 297-311
- [Me 75] Metzger, W., Gesetze des Sehens, Verlag Waldemar Kramer, Frankfurt am Main, 1975
- [Mi 68] Minsky, M., Matter, Mind and Models, in: Minsky, M., (Ed.), Semantic Information Processing, The MIT Press, MA, 1968, S. 425-432
- [Mi 86] Minsky, M., The Society of Mind, Simon and Schuster, New York, 1986
- [MiMa 97] Mignault, A./Marley, A., A Real-Time Neuronal Model of Classical Conditioning, in: Adaptive Behavior Vol. 6, No. 1, 3-61
- [Mü 97] Müller, Bernd S., Konzeptbildung und Kommunikation bei Robotern – Problemskizze und Vorschläge zum experimentellen Einstieg. Arbeitspapiere der GMD Nr. 1080, Sankt Augustin, 1997
- [Mü 98] Müller, Bernd S., Identifikation elementarer kognitiver Leistungen, (erscheint demnächst)
- [Mur 96] Murbach, A., An Integrative Computational Model for Associative Learning, in: Kognitionswissenschaft (1996) 5: 175-183
- [Ne 80] Newell, A., Physical symbol systems, in: Cognitive Science, 4 (1980) 135-183
- [Ne 90] Newell, A., Unified Theories of Cognition, Harvard University Press, Cambridge (MA), 1990
- [NeSi 76] Newell, A./Simon, H. A., Computer Science as Empirical Inquiry: Symbols and Search, in: CACM, 19, March 1976, 113-126
- [Pa 82] Papert, S., Mindstorms - Kinder, Computer und Neues Lernen, Birkhäuser, Basel 1982
- [Pal 97] Palm, G., Gehirntheorie, Neuronale Netze und assoziatives Gedächtnis, in: Northoff, G. (Hrsg.), Neuropsychatrie und Neurophilosophie, Schöningh, Paderborn, 1997, S. 105-125
- [Pe 90] Peschl, M. F., Cognitive Modelling, DUV Vieweg, 1990
- [Pen 87] Penrose, R., Minds, Machines and Mathematics, in: [BG 87], S. 259-276
- [Pi 73] Piaget, J., Einführung in die genetische Erkenntnistheorie, Suhrkamp, Frankfurt 1973
- [PM 97] Pape Møller, A., La nature préfère la symétrie, in: La recherche, 304, Décembre 1997, S. 50-55
- [PoRa 94] Posner, M. I./Raichle, M. E., Bilder des Geistes – Hirnforscher auf den Spuren des Denkens, Spektrum Akademischer Verlag, Berlin, 1994
- [PoSh 98] Poon, C.-S./Shah, J. V., Hebbian learning in parallel and modular memories, in: Biol. Cybern. 78, 79-86 (1998)
- [Pr 98] Preskill, J., Robust solutions to hard problems, in: Nature, Vol 391, 12 February 1998, S. 631-632
- [ReWo 98] Reimann, S./Wolff, J. R.: Designing Cognitive Agents, (erscheint demnächst)

- [Ri 97] Rink, J., Alice im Wunderland – Quantenrechner: Auf dem Sprung zur Realität? , in: c't 1997, Heft 3, S. 110-116
- [RiVe 91] Riegas, V./Vetter, Ch., Zur Biologie der Kognition – Ein Gespräch mit Humberto R. Maturana und Beiträge zur Diskussion seines Werkes, stw 850, Frankfurt am Main, 1991
- [RuMC 86] Rumelhart, D. E., McClelland, J. L., and the PDP Research Group, Parallel Distributed Processing, The MIT Press, Cambridge, (MA), 1986
- [RWB 43] Rosenblueth, A./Wiener, N./Bigelow, J., Behavior, Purpose and Teleology, in: Philosophy of Science, 10 (1943) 18-24
- [Si 85] Simon, H. A., The Sciences of the Artificial, Second Edition, The MIT Press, 1985
- [St 87] Stewart, I., The Problems of Mathematics, Oxford University Press, 1987
- [StKr 97] Stadler, M./Kruse, P., Zur Emergenz psychischer Qualitäten – Das psychophysische Problem im Lichte der Selbstorganisationstheorie, in: [KüKr 97], S. 134-160
- [Stru 96] Strube, G., Kognition. In: Strube, G./Becker, B./Freksa, C./Hahn, U./Opwis, K./Palm. G. (Hrsg.), Wörterbuch der Kognitionswissenschaft, Klett-Cotta, Stuttgart, 1996, S. 303-317
- [SuBa 98] Sutton, R. S./ Barto, A. G., Reinforcement Learning: An Introduction, MIT Press, Cambridge, MA, 1998
- [Tar 77] Tarski, A., Einführung in die mathematische Logik, Göttingen 1977
- [Ta 93] Tarassow, L., Symmetrie, Symmetrie! Strukturprinzipien in Natur und Technik, Spektrum Akademischer Verlag, Berlin, 1993
- [Th 98] Thorndike, E. L. Animal Intelligence – An Experimental Study of the Associative Processes in Animals, in: Psychological Review, Monogr. Suppl., Vol. II, No. 4, June, 1898, 1-109
- [Va 88] Varela, F. J., Kognitionswissenschaft - Kognitionstechnik, suhrkamp (stw 882), Frankfurt, 1988
- [Va 90] Varela, F. J., Autonomie und Autopoiese, in: Schmidt, S. J. (Hrsg.), Der Diskurs des Radikalen Konstruktivismus, suhrkamp, Frankfurt, 1990, S. 119-132
- [Ver 98] Verschure, P., Explorations in robotics and the biology of learning, in: Informatik/Informatique, No. 1, Februar 1998, S. 25-29
- [Wa 82] Wazuro, E. G., Die Lehre Pawlows von der höheren Nerventätigkeit, VVV-Verlag, Berlin, 1982
- [Wo 96] Worden, R. P., Primate Social Intelligence, in: : Cognitive Science, 20 (1996) 579-616
- [Wr 91] Wrobel, S., Die Umweltverankerung von Begriffsbildungsprozessen, in: KI – Künstliche Intelligenz, 1 (1991) 22-26