Dynamische Systeme in der KI und ihren Nachbarwissenschaften

Herbert Jaeger

Technische Fakultät, Universität Bielefeld, Postfach 100131, 33501 Bielefeld
email: herbert@techfak.uni-bielefeld.de oder herbert@arti.vub.ac.be

Mai 1995

Zusammenfassung: Diese Arbeit bietet erstens einen Überblick über bestehende Systemtheorien und einige wichtige systemtheoretische Konstrukte. Zweitens will sie ein Gefühl dafür vermitteln, welche Erklärungspotentiale systemtheoretische Modellbildungen für die KI haben. Dazu wird der Stand der Kunst bei der systemtheoretischen Modellierung intelligenzbezogener Phänomene in der KI und einigen Nachbarwissenschaften ausführlich referiert und kommentiert. Und drittens will diese Arbeit auch ein Gefühl für die praktischen und theoretischen Schwierigkeiten vermitteln, welche mit der Verwendung systemtheoretischer Techniken einhergehen.

Abstract: The paper presents, first, a crash course in system theory by providing an overview of theories and traditions, and by introducing central system-theoretic concepts. Second, it aims at demonstrating the explanatory powers of dynamical systems for AI. The state of the art in modeling phenomena related to intelligence is surveyed in AI and neighboring fields. Third, the paper explains a number of practical and theoretical impediments that oppose themselves to the application of system-theoretic methods in AI.

Inhalt

1 Einführung

Systemtheoretische Modellbildungen erfreuen sich in vielen wissenschaftlichen Disziplinen einer rasch wachsenden Beliebtheit. Sie gründet unter anderem in folgenden Punkten:

- Sehr verschiedene konkrete Systeme können mit einem einheitlichen, starken analytischen Instrumentarium untersucht werden. Diese Einheitlichkeit gründet darin, daß sich systemtheoretische Methoden vor allem anderen zeitlichen Phänomenen widmen, oder anders ausgedrückt, daß Systeme grundsätzlich als *dynamische* Systeme verstanden werden.
- Systemtheoretische Ansätze haben den Anspruch, *komplexe* Systeme mit vielen Komponenten, Wechselbeziehungen und Wirkungsebenen als Ganze verständlich zu machen.
- Exakte Analysen von Phänomenen der *Selbstorganisation* liefern neue Einsichten in teilweise seit der Antike rätselhafte Erscheinungen der Ordnungsbildung in der physikalischen, biologischen und sozialen Welt.

Außerdem ist Systemtheorie — Stichwort "Chaos" — einfach in Mode (dazu: Paslack 1989). Und nicht zuletzt geben dynamische Systeme oft schöne Computergraphiken ab.

Intelligente Systeme — Menschen, Tiere, vielleicht Roboter, vielleicht gewisse KI-Programme — sind zweifellos dynamisch, komplex und selbstorganisierend. Es liegt also nahe, systemtheoretische Methoden in der KI einzusetzen. Überdies sind systemtheoretische (enger: kybernetische) Denkweisen historisch mit den Anfängen der KI verbunden. Eine (Wieder-)Aufnahme würde einen historischen Bruch heilen, der zu einer unglücklichen Trennung der logikorientierten, fälschlich als "klassisch" angesehenen KI von der biokybernetisch-konnektionistischen Forschung geführt hat.

Gegenwärtig tauchen in der KI die ersten systemtheoretischen Ansätze auf. Ihre methodische Bedeutung ist jedoch noch nicht geklärt. In der vorliegenden Arbeit soll eine solche Klärung ansatzweise geleistet werden. Dazu ist es zunächst notwendig, eine Übersicht über die verschiedenen existierenden Systemtheorien zu geben (Kapitel 2). Etwas detaillierter werden die wichtigsten Begriffsbildungen in der Theorie dynamischer Systeme im engeren Sinne vorgestellt (Kapitel 3). Zum Stand der Kunst wird dann im 4. Kapitel referiert, wie systemtheoretische Methoden in der KI und deren Nachbardisziplinen gegenwärtig verwendet werden. Dies Kapitel enthält auch zwei knappe methodologisch Exkurse zu diskreten symbolischen Repräsentationseinheiten in kontinuierlichen Systemen und zur Natur reduktionistischer und nicht-reduktionistischer Erklärungen durch systemtheoretische Modelle. Im 5. Kapitel schließlich werden kritische Aspekte in den Vordergrund gestellt, und in Kapitel 6 wird ein kleiner Ausblick versucht.

2 Hintergrund I: Systemtheorien

Es gibt nicht "die" Systemtheorie, sondern eine Vielzahl von Traditionslinien. Sie reichen unterschiedlich weit in die Vergangenheit zurück und besitzen verschiedene Wurzeln: mathematische, biologische, physikalische, nachrichtentechnische, aber auch erkenntnistheoretische. Die verwendeten Systembegriffe sind teilweise unterschiedlich. Der gemeinsame Nenner besteht vielleicht darin, daß ein System etwas ist, das sich in der Zeit entwickelt und aus Teilsystemen oder Komponenten besteht, so daß das Ganze mehr als die Summe der Teile ist. Das ist nicht viel, und so muß man sich den einzelnen Systemtheorien genauer zuwenden. Die dann zu entdeckende Vielfalt wird aus einer kleinen alphabetischen Auswahl der insgesamt in den angesprochenen Theorien untersuchten Phänomene deutlich: Adaptation, Attraktoren, Automaten, Bifurkationen, Chaos, Differentialsysteme, Differenzierung, Emergenz, Energiefluß, Evolution, Feedback, Gestalten, hierarchische Systeme, Information, kollektive Systeme, Kompartimentierung, Konstruktivismus, Kontrolle, Lernen, Morphogenese, Musterbildung, Oszillationen, Selbstorganisation, Wachstum, Zeit.

Ein sehr empfehlenswerten ersten Einstieg liefert Kriz (1992). Eine anspruchsvollere, dabei konzise Einführung in gegenwärtige naturwissenschaftliche systemtheoretische Ideen und Methoden findet man bei Nicolis (1989). Einen Überblick zu gewinnen wird jedoch durch verschiedene Faktoren erschwert:

- Die Bezeichnungen der verschiedenen Ansätze und Traditionen werden uneinheitlich verwandt ("Kybernetik" manchmal synonym mit Regelungstechnik oder Informationstheorie, mehrere "allgemeine Systemtheorien", "Systemtheorie" kann von einem nachrichtentechnischen Methodenarsenal bis zu einer philosophischen Erkenntnistheorie für *alles* verwendet werden).
- Es gibt viele Parallelentwicklungen und zeitlich verzahnte wechselseitige Beeinflussungen. Die einzelnen "Schulen" verändern sich überdies im Laufe der Jahre.
- Systemtheoretiker sind manchmal Glaubenskämpfer. Dies führt gelegentlich zur Überbetonung von Differenzen bzw. zur aktiven Ignoranz benachbarter Ansätze. Beispiele sind etwa von Bertalanffys (1968, p.15) Darstellung der Kybernetik als Teilgebiet seiner eigenen "allgemeinen Systemtheorie", Powers' (1988) Appetenz-Vermeidungs-Konflikt gegenüber der Kybernetik, der zu seiner "kybernetischen Kontrolltheorie" führt, Hakens (1983, p. 11) Übergehen der Brüsseler Schule Prigogines bei der Besprechung oszillierender chemischer Reaktionen, oder Mandelbrots (1986) stolze Selbstbezogenheit.
- Die gegenwärtige Popularität des Phänomens "Chaos" legt einen fälschliche Gleichsetzung von "Chaostheorie" mit Systemtheorie nahe.
- Naturwissenschaftliche Systemtheorien einerseits und gesellschaftswissenschaftlich-philosophische andererseits haben inzwischen kaum noch Kontakte miteinander.

Angesichts dieser Komplikationen einerseits, und der offenbaren aktuellen Bedeutung systemtheoretischer Methoden andererseits wäre ein umfassender historischer Überblick sehr hilfreich. Leider scheint es ein solches Werk bislang nicht zu geben. Partielle historische Abrisse finden sich bei Abraham & Shaw (1983f) (zur mathematischen Geschichte kontinuierlicher dynamischer Systeme), von Bertalanffy (1968) (20er und 30er Jahre, biologisch orientiert, Querbezüge zu anderen, vorzugsweise philosophischen und psychologischen Strömungen im deutschsprachigen Raum jener Zeit; in einem später geschriebenen Appendix 1 Lesenswertes über Querbezüge zwischen Kybernetik, Kontrolltheorie und Bertalanffys Systemtheorie), Krohn, Küppers & Paslack (1987) (Ideengeschichte des Begriffs der Selbstorganisation), und Wiener (1961) (Dekade um den 2. Weltkrieg). Einen kurzen Abriß einiger Systemtheorien gibt auch von Weizsächer (1987) in einem Aufsatz, der sich der Brückenfunktion der Systemtheorien zwischen Natur- und Geisteswissenschaften widmet. Eine umfassendere historische Aufarbeitung liefert Wunsch (1985), allerdings betont aus ingenieurswissenschaftlicher Sicht und ohne Berücksichtigung der jüngeren Entwicklungen in den Naturwissenschaften und der Mathemematik, welche heute unsere Auffassung von Systemen entscheidend prägen. S.J. Schmidt (1987b) vermittelt schließlich einen weitgehenden Überblick über die nicht-naturwissenschaftlichen Systemtheorien. Diese Arbeit ist jedoch nicht historisch, sondern nach Ideen und Anwendungen geordnet.

An dieser Stelle kann ein Gesamtüberblick nicht versucht werden. Ersatzweise sollen einige prominente Systemtheorien in Stichpunkten vorgestellt werden.

von Bertalanffys "Allgemeine Systemtheorie" (von Bertalanffy 1968), schon in den 20er Jahren begonnen, ist nach Ansicht des Autors die erste und originale Systemtheorie. Sie ist als Methode der Biologie konzipiert. Erst nach dem 2. Weltkrieg kommt es zum Durchbruch mit der Gründung einer "Society for General System Theory" am Centre for Advanced Studies in Palo Alto 1954 (später "Society for General System Research"). Einige Charakteristika dieser Systemtheorie sind im folgenden aufgezählt:

- Definition: System ist eine Menge von Elementen, die in Relationen zueinander stehen, welche das Verhalten der Elemente beeinflussen. Dadurch ist das System mehr als die Summe der Elemente.
- Mathematisch werden die Elemente durch kontinuierliche Variable, die Relationen durch Differentialgleichungen formalisiert, genauso wie in vielen jüngeren naturwissenschaftlichen Systemtheorien (Kybernetik, Synergetik).
- Im Gegensatz etwa zu Wiener (Kybernetik) oder Haken (Synergetik) entwickelt von Bertalanffy keine eigenen mathematischen Techniken.
- Mutet philosophischer an als andere naturwissenschaftliche Systemtheorien. Diskutiert werden z.B. Fragen nach Finalität/Kausialiät, Ganzheit/Gestalt, Relativität von Beschreibungskategorien.
- Mutet biologischer an als andere naturwissenschaftliche Systemtheorien, indem Fragen des Wachstums, Morphogenese, Gestaltbildung, etc., diskutiert werden.

Aus heutiger Sicht und für die KI ist von Bertalanffys "Allgemeine Systemtheorie" nicht so sehr wegen des mathematisch-formalen Gehalts von Interesse (von Bertalanffy 1968 ist aber ein didaktisch vorzüglicher Einstieg). Mathematisch sind die jüngeren, in manchem vergleichbaren Ansätze der Kybernetik und der Synergetik wesentlich weiter entwickelt. Die Lektüre von v. Bertalanffy lohnt gleichwohl, weil hier noch ein im Vergleich zu den genannten späteren Entwicklungen phänomenal besonders reichhaltiger und lebendiger Systembegriff vorliegt. Er umfaßt Aspekte wie funktionale und strukturelle Hierarchien, Wachstum, progressive Mechanisierung (d.h. Automatisierung von Prozessen, Reflexbildung), progressive Segregation (d.h. Ausdifferenzierung eines Systems in immer unabhängigere Teilsysteme), Zentralisierung u.v.a.m. (alle in v. Bertalanffy 1986). Dank ihrer biologischen Wurzeln steht diese Systemtheorie näher als andere an dem für die KI so wichtigen Thema lernender, sich differenzierender Systeme.

Die **Kybernetik** ist in der Zeit um den 2. Weltkrieg im engen intellektuellen Austausch von Wissenschaftlern aus Mathematik, Neurophysiologie und Psychiatrie, Regelungstechnik und anderen Gebieten entstanden (Beschreibung dieser Epoche in Wiener 1961[2]). Der erste Anlaß waren Fragen der nichtlinearen stochastischen Zeitreihenvorhersage bei der Flugabwehrgeschützsteuerung, denen sich Norbert Wiener eine Zeitlang widmete. Als zentrale Figuren kann man wohl den Mathematiker Norbert Wiener, den Ingenieur Julian H. Bigelow und den Neurologen Arturo Rosenblueth ansehen. Die Kybernetik wurde aber auch mit geformt durch von Neumann (Ergodentheorie, digitale Informationsverarbeitung), McCulloch und Pitts (erste mustererkennende neuronale Netzwerke, nach heutiger Terminologie), Lewin (Psychologie), Bateson und Mead (Anthropologie), sowie viele andere. Bei einer solch vielseitigen Ahnenschaft ist klar, daß es nicht *die* Kybernetik gibt. Tatsächlich wird bis in die heutige Zeit der Begriff oftmals fast nach Belieben für jedwede Forschungen gebraucht, die etwas mit Regelung, Informationsverarbeitung, biologischen Systemen oder anderen Facetten der Forschungen in der Gründerzeit zu tun haben. Im folgenden werden einige Merkmale der "Basis-Kybernetik", wie sie in Wiener (1961[2]) niedergelegt ist, angegeben:

- Es werden sowohl kontinuierliche als auch diskrete Systeme behandelt. Dabei werden Techniken der mathematischen Analysis mit solchen der Informationstheorie verknüpft.
- Vielerlei mathematische Techniken finden Verwendung und werden teilweise weiterentwickelt, u.a. Differential- und Integralgleichungen, Ergodentheorie, stochastische DGL, Automatentheorie, Informationstheorie. Laut Wiener sollte die Kybernetik letztlich in mathematischer Logik begründet werden. Dies wird aber nicht durchgeführt.
- Theoretische Grundaufgaben sind Fragen der Zeitreihenvorhersage, der Stabilität und Oszillation in Systemen mit Feedback, und der Mustererkennung.
- Ein wichtiges Anwendungsfeld: biologische Steuerung und Informationsverarbeitung.

Aus heutiger Sicht ist die Kybernetik eine historische Erscheinung. Seit den 40er und 50er Jahren sind die in ihr versammelten Themen wieder mehr in einzelne Disziplinen auseinandergefallen: etwa die biologische Kybernetik, die Automatentheorie, die Informationstheorie, die ingenieurswissenschaftliche Kontrolltheorie, die Mustererkennung in der Informatik, und andere. Der Begriff "Kybernetik" hat jedoch überlebt, wenngleich auf Kosten einer nahezu beliebigen Verwendbarkeit.

Für die KI ist die klassische Epoche der Kybernetik von großem Interesse. Durch ihre ursprüngliche Verbindung von diskreten Automaten und diskreter Informationstheorie mit neuronal inspirierten, kontinuierlichen Musterdetektoren könnte eine Neubetrachtung der klassischen Kybernetik zwischenzeitlich verschüttete Wege zur Reintegration der diskreten symbolischen KI mit dem Konnektionismus wieder öffnen. Ein solcher Weg könnte in einer vereinheitlichenden informationstheoretischen Fassung beider Gebiete liegen.

Eine weitgehend selbständige **ingenieurswissenschaftliche Systemtheorie** (mit dem prominenten Teilgebiet der **Kontrolltheorie**) hat sich seit Beginn des Jahrhunderts in der Nachrichten- und Regelungstechnik entwickelt (geschichtlicher Abriß mit formaler Einführung zentraler Techniken: Wunsch 1985). Es ging zunächst um die Analyse der Eigenschaften von (elektrischen) Systemen bzw. um die Synthese solcher Systeme bei vorgegebenen Input-Output-Verhalten. Der Begriff "Systemtheorie" wird zuerst 1949 von Küpfmüller (referiert nach Wunsch 1985) verwendet. Küpfmüller behandelt komplexe elektrische Netzwerke, die aus linearen Übertragungsgliedern zusammengesetzt sind. Hierbei geht er von den mathematischen Übertragungscharakteristiken der Elemente aus und sieht von deren physikalischer Verwirklung ab (black-box-Methode).

Eine zentrale Aufgabenstellung innerhalb der ingenieurswissenschaftlichen Systemtheorie betrifft die Regelung eines Systems, d.h. die Einhaltung bzw. Wiederherstellung gewünschter Systemzustände durch geeigneten Input (Kontrolltheorie, Regelungstheorie). In den 60er Jahren wird von Kalman und anderen eine Grundaufgabe der Kontrolle linearer Systeme gelöst. Sie besteht in der optimalen Schätzung des Zustandes eines stochastischen Systemes, über das nur stochastisch gestörte und unvollständige Beobachtungsinformation vorliegt. Dies Schätzverfahren, der sog. Kalman-Filter, und seine Verallgemeinerungen bilden den Grundbestand der heutigen Kontrolltheorie (Querschnitt von Einzeldarstellungen in Zadeh & Polak 1969, schönes Lehr- und Handbuch Stengel 1986). Die ingenieurswissenschaftliche Systemtheorie besitzt eine eigenständige, anwendungsorientierte Tradition, die von "akademischeren" Systemtheorien kaum rezipiert wird. Was die in der Kontrolltheorie verwendeten mathematischen Techniken (Filterung und Prädiktion stochastischer Prozesse) betrifft, so bestehen Überschneidungen mit der klassischen Kybernetik.

Die ingenieurswissenschaftliche systemtheoretische Tradition ist keine einheitliche Schule, sondern eher ein ordnender Gesichtspunkt für eine Vielzahl von Techniken. Diese sind reichlich oberflächlich durch die folgenden beiden Merkmale charakterisiert:

- Systeme werden als Einheiten mit einem Input-Output-Verhalten verstanden. Dies steht im Gegensatz zu fast allen anderen Systemtheorien, wo allein der interne Zustand eines System betrachtet wird.
- Im Gegensatz zu anderen naturwissenschaftlichen Systemtheorien geht es nicht nur um die *Analyse* von Systemen, sondern mindestens ebenso sehr um den *Entwurf* von Systemen mit vorgegebenen gewünschten Eigenschaften, und um die Regelung (Kontrolle) von technischen Systemen.

In der Robotik und der Mustererkennung werden Techniken der Kontrolltheorie standardmäßig verwendet. Will man diese Disziplinen zur KI zählen, so ist die Kontrolltheorie also bereits in der KI etabliert. Die Kontrolltheorie könnte jedoch in einem viel umfassenderen Sinne für die KI interessant sein. Aus der Wahrnehmung eines Kontrolltheoretikers liest sich das so (Stengel 1986, p.4): *"There is a natural relationship between stochastic optimal* *control and* [...] *artificial intelligence. AI algorithms are applied to the interpretation of data, system monitoring and identification, prediction, planning, and design. They may use "production systems" to model dynamic interactions predicated on large data bases, as well as optimization techniques for decision making under uncertainty. Stochastic optimal control theory provides tools for solving many of these AI problems."* Die in diesem Zitat herausgestellte prinzipielle Homologie zwischen Grundaufgaben der Kontrolltheorie und der KI ist einleuchtend. Tatsächlich ist es bisher jedoch nicht zu einer Integration von kontrolltheoretischen Methoden in die in diesem Zitat angesprochenen klassischen KI-Anwendungen gekommen.

Zadehs "allgemeine Systemtheorie" ist historisch aus der ingenieurswissenschaftlichen Tradition erwachsen. Inhaltlich ist sie aber eine selbständige, abstrakte, nahezu "reine" mathematischen Theorie. Sie wurde in den frühen 60ern von L.A. Zadeh entwickelt (Einführung in Zadeh 1969). Ihre Kennzeichen sind:

- Ein System wird mengentheoretisch als Familie von Paaren von Zeitfunktionen formalisiert. Zustände ergeben sich formal als gewisse Teilmengen dieser Familie.
- Der Systembegriff ist so abstrakt, daß sich die in anderen mathematisierten Systemtheorien betrachteten Systeme (z.B. Differentialsysteme, Automaten) als Spezialfälle interpretieren lassen.

Die mengentheoretisch abstrakte Spielart von Systemtheorie hat, wie es scheint, wenig Einfluß gewinnen können. Sie ist möglicherweise zu abstrakt, um noch interessante Aussagen zu ergeben. Dennoch könnte sie für die KI relevant werden. Wie am Ende dieser Studie ausgeführt wird, könnte die KI genötigt sein, neuartige Systemtypen zu erfinden. Die mengentheoretische allgemeine Systemtheorie könnte dann helfen, diese neuartigen Systeme mit schon bekannten zu vergleichen, z.B. indem sie hilft, kanonische Zustandsbegriffe zu definieren.

Die vom Neurobiologen Humberto R. Maturana und dessen Schüler Francisco J. Varela eingeführten **autopoietischen Systeme** sind ursprünglich ein Ansatz, Prinzipien lebendiger Systeme zu erklären. Maturana und Varela selbst bezogen bald Phänomene der Kognition und des Bewußtseins in ihre Untersuchungen ein (Einführung in Maturana & Varela 1984). Obwohl die Autoren mit konkreten physiologischen Beispielen nicht sparen, ist die Grundhaltung eher philosophisch und erkenntnistheoretisch als naturwissenschaftlich. Der Einfluß der Theorie autopoietischer Systeme in den Sozial- und Geisteswissenschaften, aber auch in der Psychiatrie und der Literaturwissenschaft ist kaum zu überschätzen (Beispiele in S.J. Schmidt 1987b). Der Begriff "Systemtheorie" wird in solchen Gebieten oftmals synonym mit der von Maturana und Varela ausgehenden Tradition verwandt (z.B. Willke 1987). Die erkenntnistheoretische Grundthese, daß nämlich lebendige Systeme die äußere Wirklichkeit nicht intern abbilden, sondern je eine eigene, wesentlich durch das System bestimmte Wirklichkeit konstruieren, hat zu einer eigenen erkenntnistheoretischen Richtung in der Philosophie geführt, dem (radikalen) Konstruktivismus (Reader: S.J. Schmidt 1987a).

Die von Maturana und Varela gepflegte philosophische, ja manchmal visionäre Art, Systemtheorie zu betreiben, ist weit entfernt von den eher "harten" naturwissenschaftlichen Systemtheorien, die in dieser Studie sonst im Vordergrund stehen. Dennoch gibt es eine ganze Reihe von Forschern, die sowohl von der philosophischen als auch von den "hart" naturwissenschaftlichen Denkrichtungen inspiriert werden (wie etwa von Foerster, Roth, von Glasersfeld — Aufsätze dieser Autoren in Schmidt S.J. 1987a).

Einige Merkmale der Theorie autopoietischer Systeme sind:

- Der Kernbegriff "Autopoiesis" ist schwer präzise zu fassen. Er betrifft die charakteristische Eigenschaft von Lebewesen, daß sie sich "*buchstäblich — andauernd selbst erzeugen.* [...]

Im wesentlichen ist diese Organisation durch gewisse Relationen gegeben, die ... auf der zellulären Ebene noch leicht zu verstehen sein werden." (Maturana & Varela 1984 p.50f)

- Maturana und Varela beschreiben, von der molekularen Ebene ausgehend, immer höhere Organisationsstufen des Lebens, bis zum menschlichen Bewußtsein. Der Begriff der Autopoiese wird dabei zunehmend abstrakt (und zunehmend schwer verständlich).
- Autopoiese ist gekoppelt mit Selbstreferenz und Abgeschlossenheit. Aus dem Bild einer Zelle, die sich im großen und ganzen nach ihren eigenen Gesetzmäßigkeiten formt und durch eine Membran nach außen abgeschlossen ist, ergibt sich das Bild eines Nervensystemes und eines menschlichen Geistes, deren Prozesse ebenfalls weitgehend unabhängig von der äußeren Wirklichkeit sind.

Bedingt durch ihren philosophischen Charakter kann die Theorie autopoietischer Systeme für die KI naturgemäß keine konkreten Hilfen beim Design von Systemen geben, sondern wird eher in methodologischen Diskussionen zum Tragen kommen. In der KI sind aber autopoietische Systeme bisher kaum thematisiert. Der einzige mir bekannte Fall einer eingehenderen Rezeption ist das Buch von Winograd & Flores (1986). Die Theorie autopoietischer Systeme könnte jedoch fruchtbar in die epistemologische Diskussion um die Perspektive der "Situated Action" eingebracht werden. Für diese Perspektive ist ein mehr oder weniger ausgeprägte Anti-Repräsentationalismus und die Betonung des Systemcharakters einer Agent/Umwelt-Einheit kennzeichnend (z.B. Suchman 1987, Greeno & Moore 1993, Clancey 1993), genauso wie für die Position von Maturana und Varela. In verstreuten Quellen sind dahingehende Bemerkungen zu finden (z.B. Smithers 1994, Beer 1995), die aber bislang noch nicht zu einer systematischen Diskussion geführt haben.

Evolutionstheorie und **Populationsdynamik** (Wilson & Bossert 1973) sind zwar weder eine einheitliche Disziplin noch wohl im Selbstverständnis der meisten dort tätigen Forscher systemtheoretische Traditionen, dennoch können sie hier mit guter Berechtigung aufgeführt werden. Die Populationsdynamik liefert einige Standardbeispiele dynamischer Systeme (Räuber-Beute-Systeme, Vermehrung unter begrenzten Ressourcen), die mit typisch systemtheoretischen Mitteln etwa auf Periodizitäten, Bifurkationen und Chaos (siehe Kapitel 3) untersucht werden (vgl. Haken 1983 p.14). Die von Eigen & Schuster (1977f) bei der Untersuchung des Hyperzyklus verwendeten mathematischen Methoden stammen aus der differentialgeometrischen Theorie dynamischer Systeme. Die Verwendung solcher mit anderen Systemtheorien geteilten Techniken charakterisieren Evolutionstheorien als Systemtheorien. Das Gebiet hat jedoch eine betont eigene Ausprägung:

- Es werden sehr verschieden lange (und sehr lange) Zeiträume berücksichtigt (Beispiel: Erklärung von Artenzusammensetzung und dynamischer Stabilitäten in Distelkopf-Kleinbiotopen vor dem Hintergrund dreier um Größenordnungen getrennter Zeitskalen in Zwölfer 1986).
- Räumliche, relativ langsam veränderliche Strukturen werden als Bedingungen für relativ schnelle Artendifferenzierungsprozesse untersucht.
- Ökosysteme als (ein) Gegenstand der Untersuchung sind ausgeprägt offene Systeme, deren Modellierung außerdem mit einem besonderen Maß an Zufall fertigwerden muß. Das führt zu eigenen Modifikationen des Systembegriffs (Bemerkungen dazu in Zwölfer 1986).
- Es werden nicht nur sich zeitlich verändernde Systeme, sondern *Mechanismen zeitlicher Veränderung* selbst (Evolutionsmechanismen) untersucht. Es wird sogar die Evolution von Evolutionsmechanismen erforscht (Wagner 1986), z.B. verschiedene Varianten von Sexualität. Ferner werden Evolutionsmechanismen auf verschiedenen Ebenen betrachtet, von den Genen (besonders in der sog. Soziobiologie betont; Reader: Clutton-Brock & Harvey 1978) über die Arten bis zu "makroevolutionären" Artensystemen.

Insgesamt ist die Evolutionstheorie (genauer, die Evolutionstheorien, da es ausgeprägte Schulenbildungen gibt) mit einem ungeheuren Phänomenenreichtum befaßt, für die es wohl keine einfache, universale Theorie oder ein geschlossenes mathematisches Instrumentarium geben kann. Einen lebendigen Eindruck dieser Vielfalt vermittelt die Sammlung populärer Aufsätze von Gould (1979) und der methodologische Aufsatz von Wagner (1986).

Für die Informatik und die KI interessant sind einige Techniken zur evolutionären Optimierung (Handbuch: Schwefel 1977), die auf einer einfachen Grundversion der Darwin'schen Prinzipien von Mutation und Selektion beruhen. Die beiden wichtigsten Ansätze sind genetische Algorithmen, die zumeist mit Classifier-Systemen verwendet werden (Klassiker: Holland 1975, Handbuch: Goldberg 1989, Einführung: Holland 1986), und Evolutionsstrategien zur stochastischen Optimierung (Rechenberg 1973, referiert nach Wagner 1986). Die im Vergleich zu biologischen genetischen Systemen sehr einfachen formalen Classifier-Systeme können gut als dynamische Systeme formalisiert und mit bekannten Methoden analysiert werden (Forrest & Miller 1990). Sie haben inzwischen in der Informatik eine weite Verbreitung gefunden.

Physikalisch-mathematische Theorien der Selbstorganisation sind verantwortlich für den in den enormen Aufschwung systemtheoretischer Methoden primär innerhalb und sekundär außerhalb der Naturwissenschaften. Zu nennen sind hier vor allem die eben schon angesprochene Theorie des Hyperzyklus von Eigen und Schuster (1977), die Theorie der Selbstorganisation fern vom thermodynamischen Gleichgewicht der "Brüsseler Schule" (Prigogine 1980), und die Synergetik von Haken (populäre Einführung in Haken & Wunderlich 1986, Standardreferenz Haken 1983). Die gemeinsame Fragestellung ist die Entstehung von Ordnung in Systemen, die aus (sehr) vielen gekoppelten Teilsystemen bestehen. Viele neuartige Einsichten und kraftvolle, allgemeine mathematische Techniken sind erarbeitet worden. Es zeigt sich, daß hinter dem intuitiven Begriff der Selbstorganisation eine große Phänomenvielfalt liegt. Die Synergetik und die Brüsseler Schule sind keine genau umrissenenen formale Einzeltheorien, sondern Sammlungen analytischer Methoden zur Modellierung komplexer Systeme. Diese mathematischen Techniken sind teils Fortschreibungen aus dem kanonischen kybernetischen und differentialgeometrischen Methodenarsenal, teils von Haken bzw. Prigogine und ihren Mitstreitern selbst entwickelte speziellere Methoden. Kennzeichnend sind:

- Ein durchgängiges Thema sind Entwicklungspfade über Bifurkationen (Prigogine 1980, Kap. V, VI), spezieller auch "Wege ins Chaos" (Haken 1983 p. 264ff)
- Ein weiteres wichtiges Thema ist Selbstorganisation in kollektiven Systemen mit vielen Freiheitsgraden. Die Synergetik trägt hier durch originäre Techniken zur Behandlung von sog. Versklavungsphänomenen (vgl. Kapitel 3) wesentliche Einsichten bei.
- Entwickelt sich ein System entlängs einer Reihe von Bifurkationen, so ist zum Verständnis eines späteren Systemzustands die Kenntnis von dessen *Geschichte* nötig. An den Bifurkationspunkten bestimmen mikroskopische zufällige stochastische Fluktuationen die makroskopische Entwicklung. Dieses Prinzip der "order through fluctuations" wird vor allem von Prigogine (1980) thematisiert.
- Im Vergleich zur verwandten, rein mathematischen, differentialgeometrischen Theorie dynamischer Systeme (s.u.) steht die rechnerisch-praktische Beschreibung empirischer Phänomene im Vordergrund: ein Verhältnis etwa wie zwischen reiner Kunst und Handwerk.

In den Nachbargebieten der KI ist besonders die Synergetik als "die" Systemtheorie rezipiert worden. Das hängt sicher damit zusammen, daß Hermann Haken seine Ideen mit viel Charisma und Einsatz außerhalb der Naturwissenschaften verbreitet. Synergetische Methoden haben ihren Eingang in Disziplinen wie z.B. die Wirtschaftswissenschaften (Ulrich & Probst 1984) oder die Psychologie (Beispiele in Kriz 1992, 150ff) gefunden.

Die **differentialgeometrische Theorie dynamischer Systeme** in der Mathematik beschäftigt sich mit im wesentlichen mit der qualitativen Analyse der Lösungsmengen von (zeitlichen) Differentialgleichungssystemen. Sie geht historisch auf die mathematische Behandlung von Problemen der Himmelsmechanik zurück und enthält große Teile des Bestandes der klassischen Differentialgeometrie. In jüngerer Zeit hat eine eigenständige, spezialisierte Forschung eingesetzt, die sich um die strenge mathematische Durchdringung von Phänomenen kümmert (Attraktoren, Chaos, Bifurkationen u.a.), die oft auch in den naturwissenschaftlichen Systemtheorien thematisiert werden. Im folgenden Kapitel werden diese eingehender beschrieben, so daß hier auf eine Charakterisierung verzichtet werden kann. Eine sehr schöne, ausführliche und überaus reich bebilderte Einführung bieten Abraham & Shaw (1983f). Dieses erstaunliche Werk setzt keine Vorkenntnisse voraus. Eine konzise, für mathematisch nicht speziell vorgebildete Naturwissenschaftler zugängliche Darstellung zentraler Konstrukte findet sich bei Kelso et al. (1993). Ein einführendes mathematisches Lehrbuch ist Arrowsmith & Place (1992), und Arrowsmith & Place (1990) ist ein umfassendes Lehrbuch mit Handbuchcharakter, das an vielen Stellen an die Forschung heranführt.

In den hier angeführten Systemtheorien werden dynamischen Systeme zumeist in **kontinuierlichen** Variablen beschrieben (Ausnahmen sind in der Kybernetik und der nachrichtentechnischen Systemtheorie zu finden). Typischerweise werden hierbei dynamische Systeme durch Differentialgleichungen oder Diffeomorphismen (vgl. Kap. 3) spezifiziert. So lohnt es sich vielleicht, abschließend ausdrücklich festzustellen, daß dynamische Systeme genausogut auch für **diskrete** Beschreibungsgrößen definiert werden können. Die wichtigsten Typen sind hier wohl (endliche) Automaten und Zellularräume (engl. *celluluar automata*). Erstere sind Informatikern natürlich im Rahmen der Theorie formaler Sprachen und der Komplexitätstheorie wohlbekannt, und letztere sind ein mathematisch einfaches Modell von Parallelrechnern. Nicht vielen dürfte noch bewußt sein, daß (endliche) Automaten historisch mit der Informationstheorie und der Kybernetik verbunden sind (Einführungen, die noch in dieser Tradition stehen, in Gill 1969 und Carlyle 1969); das Wort "Kybernetik" wurde bis in die 70er Jahre auch im Sinne von "digitaler Informationsverarbeitung" verwendet. Dieser historische Zusammenhang hat sich aufgelöst. In der KI zumal dürften Automatenmodelle der traditionellen symbolischen Richtung zugerechnet werden und damit eher als Gegensatz zu systemtheoretisch orientierten Neuansätzen empfunden werden.

Tatsächlich gibt es jedoch verschiedene Forschungen, in denen endliche Automaten, bzw. diskrete Prozesse, die Symbolsequenzen produzieren, im echt systemtheoretischen Sinne untersucht werden. Eine dieser Forschungslinien, die hautpsächlich von theoretischen Physikern betrieben wird, untersucht solche Prozesse mit Methoden der statistischen Mechanik bzw., was im großen und ganzen gleichbedeutend ist, mit Methoden der Informationstheorie (z.B. J.S. Nicolis & Tsuda 1988, Crutchfield & Young 1990, Ebeling & G. Nicolis 1992). Hierbei können aus Häufigkeitsverteilungen von Teilsequenzen in beobachteten Symbolsequenzen stringente Rückschlüsse auf die Natur des generierenden Prozesses (chaotisch, Markoff'sch etc.) geschlossen werden. Thomas (1977) hat unter der Bezeichnung "Kinetische Logik" endliche Automaten als diskrete Approximationen kontinuierlicher Systeme aufgefaßt und z.B. zur Modellierung von DNA-Regulationsmodellen in der Biologie angewendet. Thomas et al. (1977) geben eine interessante vergleichende Diskussion von differentialgeometrischen, stochastischen und boole'schen Systemmodellen. Zellularräume werden als Modell selbstorganisierender, räumlich ausgedehnter Systeme von Physikern verwendet und von Mathematikern untersucht. Dies ist inzwischen ein ausgedehntes Forschungsfeld (Wolfram 1983, Sammelband Wolfram 1986). In die KI haben bisher weder (im heutigen Sinne) systemtheoretische Automatentheorien noch Zellularräume Eingang gefunden. Das gilt auch für die Nachbargebiete der KI, so daß diese Ansätze in diesem Übersichtswerk nur am Rande behandelt werden.

Die Vorstellung wichtiger Systemtheorien sei durch eine kleine Ahnentafel abgerundet, die naturgemäß unvollständig ist (Tabelle 1). Einige Forscher in dieser Tabelle müßten richtigerweise in mehrere Spalten eingetragen werden (z.B. Wiener noch bei Mathematik und nicht-Naturwissenschaft, Ashby noch bei Chemie und Biologie, Maturana & Varela noch bei Biologie/Chemie).

Mathematik	**Physik**	**Technik**	**Chemie und Biologie**	**nicht Naturwissenschaft**
Jules Henri Poincaré (1854-1912) Differentialgeometrie				Jakob Johan von Uexküll (1864-1944) Umwelt und Innenwelt der Tiere 1909 Theoretische Biologie (20er)
Aleksandr Michail Liapunov (1857-1918) Stabilitätstheorie		John William Strutt (Lord Rayleigh) (1842-1919) nichtlineare Oszillationen		
Georg David Birkhoff (1884-1944) Ergodentheorie, Phasenraum		Balthasar van der Pol (1889-1959) elektrische Schwingungen	Ludwig von Bertalanffy Allgemeine Systemtheorie (~ 1925)	
W.R. Ashby (1903-1972) diskrete Markov-Prozesse (Kybernetik)		Claude E. Shannon Informationstheorie (40er)		
John von Neumann (1903-1957) zelluläre Automaten ~ 1950 (Kybernetik), Ergodentheorie, Spieltheorie		Norbert Wiener (1894-1964) Kybernetik 1949		
K. Itô Stochastische DGL (50er)		K. Küpfmüller (1897-1977) Systemtheorie der Signalübertragung (1949)	Heinz von Foerster Selbstorganisation (1960) Soziale Systeme, Erkenntnistheorie	
L.A. Zadeh abstraktes System (1963)	Edward N. Lorenz (Meteorologie) chaotischer Attraktor 1961			
			Ilja Prigogine (*1917) irreversible Thermodynamik (50er) dissipative Systeme, Selbstor- ganisation (~ 70er)	
René Thom Katastrophentheorie (1966)	Hermann Haken (*1927) Lasertheorie ~ 1962 Synergetik ab 1970			
Benoit B. Mandelbrot (*1924) Fraktale Geometrie 1982			William T. Powers Kybernetische Kontrolltheorie(60er, 70er)	Humberto R. Matuarana (*1928) Francisco J. Varela (*1946) Autopoietische Systeme (ab ~ 1970) Konstruktivismus
			Manfred Eigen (*1927) Hyperzyklus, molekulare Evolution (~ 70er)	Niklas Luhmann Soziale Systeme (ab ~ 1975)

Tabelle 1: Eine "Ahnentafel" systemtheoretischer Forscher.

3 Hintergrund II: Wichtige Phänomene bei kontinuierlichen dynamischen Systemen

In diesem Kapitel werden die wichtigsten Begriffe und Phänomene kontinuierlicher dynamischer Systeme (KDS) vorgestellt. Für die KI sind KDS gegenwärtig die wichtigsten Typen dynamischer Systeme; sie sind fast immer gemeint, wenn in der Literatur von "dynamischen Systemen" die Rede ist. In der Mathematik werden sie von einem inzwischen weitgehend spezialisierten Zweig der Differentialgeometrie behandelt, der "Theorie dynamischer Systeme" (*dynamical systems theory*). KDS sind auch der Hauptgegenstand der formalen Techniken in den naturwissenschaftlichen Systemtheorien, dort dann oft mit einer stochastischen Komponente.

Vorweg eine terminologische Bemerkung: unter "kontinuierlichen Systemen" verstehe ich solche, deren *Zustands*variable kontinuierlich sind, und unter "diskreten Systemen" solche mit diskreten Zustandsgrößen. In der Literatur findet man alternativ genauso häufig auch das Attribut "kontinuierlich" für Systeme mit kontinuierlicher *Zeit*, und entsprechend von "diskret" für Systeme mit diskreter Zeit.

Die differentialgeometrische Theorie dynamischer Systeme und die naturwissenschaftlichen Methoden der Synergetik etc. unterscheiden sich durchaus in Themenstellung und Strenge. Letztere sind am Verständnis realer Phänomene interessiert und benutzen dafür an mathematischen Techniken alles, was zur Verfügung steht, aber nur in der Strenge und Allgemeinheit, die gerade noch ausreicht. Die Synergetik z.B. wirkt daher mathematisch eklektischer und pragmatischer als die differentialgeometrische Disziplin. Beiden Herangehensweisen ist aber ein großer Grundbestand an Themen und Techniken gemein. Vor allem dieser gemeinsame Bereich wird auch von außerhalb der Mathematik bzw. der Naturwissenschaften rezipiert. Daher wird in diesem Kapitel von den disziplinären Akzentuierungen abgesehen und einfach "die" KDS vorgestellt werden. Das geschieht relativ knapp. Es gibt anderswo gute Einführungen (nahezu populär: Haken & Wunderlich 1986, für den durchschnittlich mathematisch vorgebildeten Naturwissenschaftler in Kelso et al. 1993 — mathematische Akzentsetzung — und das erste Kapitel in Haken 1983 — physikalische Akzentsetzung — , besonders umfassend, aber auch knapp und anspruchsvoll in Nicolis 1989).

Ein KDS besteht aus einem kontinuierlichen *Zustandsraum* (synonym: *Phasenraum*), dem eine Dynamik aufgeprägt ist, d.h. eine Gesetzmäßigkeit, welche festlegt, wie im Laufe einer kontinuierlich fließenden Zeit die Zustände aufeinander folgen. Ein solcher Zustandsraum ist typischerweise der $\mathbb{R}^n$ bzw. eine differenzierbare reelle Mannigfaltigkeit (seltener auch komplexwertige Zustandsräume), mit Dimensionen $x_1, ..., x_n$. Ein Zustand ist also durch einen n-dimensionalen Vektor $(x_1, ..., x_n)$ gegeben, in Kurzschreibweise $\mathbf{x}$.

Betrachtet man ein KDS als Modell für ein reales System, so repräsentiert jede Dimension eine "Meßgröße". Ferner findet man auch noch folgende Deutungen bzw. Bezeichnungen der Dimensionen in der (synergetisch beeinflußten) Literatur: Komponenten, Teilsysteme, Freiheitsgrade, Moden, Zustandsgrößen, Systemvariable. Diese inhaltlichen Aspekte sind für die mathematische Theorie jedoch irrelevant.

Die zeitliche Dynamik, mit der die Zustände ineinander übergehen, kann durch verschiedene mathematische Techniken spezifiziert werden. Die wichtigsten sind Diffeomorphismen, Ströme und Differentialgleichungen (DGL).

A. *Diffeomorphismen* **f** sind (differenzierbare) Abbildungen des Zustandsraumes in sich (Arrowsmith & Place 1990, 5ff). Ein einfaches Beispiel ist der Diffeomorphismus $\mathbf{f}: x \to x^2$ auf dem Zustandsraum $\mathbb{R}$.

Die Anwendung von **f** auf einen Zustand **x** kann als ein diskreter Zeitschritt gedeutet werden. Wenn man die Zeit durch Indizes *t* markiert, kann man auch schreiben: $\mathbf{x}_{t+1} = \mathbf{f}(\mathbf{x}_t)$. Da $\mathbf{x} = (x_1, ..., x_n)$ ein Vektor ist, ist damit gleichwertig die Angabe von n Gleichungen für die einzelnen Komponenten. Berühmtes Beispiel: die Hénon-Abbildung

$$x_{t+1} = x_t \cos\alpha - y_t \sin\alpha + x_t^2 \sin\alpha$$
$$y_{t+1} = x_t \sin\alpha + y_t \cos\alpha - x_t^2 \cos\alpha$$

Diese viel untersuchte Abbildung ist bemerkenswert durch die Reichhaltigkeit der in ihr verborgenen Phänomene (Arrowsmith & Place 1990, 56ff). Die Hénon-Abbildung entspricht zwar einer einfachen Idee: sie ist inhaltstreu und im linearen Anteil eine reine Rotation. Dennoch führt sie zu Phänonemen wie *"... Inselketten, chaotische Bahnen und ihre Wiederholung auf allen Skalen. ...* [ein] *Bild von immenser Komplexität, das noch lange nicht voll verstanden ist"* (Arrowsmith & Place 1990, 56). Solche Situationen, wo schon die Untersuchung "einfacher" KDS an die Grenzen der derzeitig verfügbaren Analysetechniken stößt, sind typisch für die Mathematik von KDS. Grob gesagt, braucht es zumeist nur einen nichtlinearen Anteil der Dynamik, um viele derzeit noch unverstandene dynamische Phänomene hervorzubringen.

Ein Folge ... , $\mathbf{x}_{t-1}$, $\mathbf{x}_t$, $\mathbf{x}_{t+1}$, $\mathbf{x}_{t+2}$, ... von Zuständen ist eine *Trajektorie* (Bahn) eines Diffeomorphismus. Zwei Trajektorien des Diffeomorphismus $\mathbf{f}: x \to x^2$ auf $\mathbb{R}$ wären etwa 1, 1, 1, ... und 2, 4, 8, ... Man darf sich hier nicht verwirren lassen: der Zustandsraum ist kontinuierlich, der Diffeomorphismus sogar differenzierbar, dennoch sind die Trajektorien diskret.

B. *Ströme* erhält man, wenn man statt eines einzelnen Diffeomorphismus eine ganze Familie $(\varphi_t)_{t \in \mathbb{R}}$ von solchen nimmt. Die einzelne Abbildung φ_t beschreibt dabei, wie sich das System in einem Zeitraum der Länge t verändert (Arrowsmith & Place 1990, 14ff). Man verlangt $\varphi_0 = \mathrm{id}$ (d.h., wenn keine Zeit vergeht, ändert sich das System auch nicht), und $\varphi_t\varphi_s = \varphi_{t+s}$ (d.h., wenn ich das System sich erst um die Zeit s, dann um die Zeit t entwickeln lasse, dann entwickelt es sich insgesamt um die Zeit $s+t$). Ferner verlangt man, daß $(\varphi_t)_{t \in \mathbb{R}}$ in t stetig differenzierbar ist (d.h., das System verändert sich in der Zeit "glatt").

Ein einfaches Beispiel ist der zweidimensionale Strom $\varphi_t((x, y)) := (x + t, y + 2t)$. Er beschreibt eine Translation des ganzen Raumes mit konstanter Geschwindigkeit 1 in x-Richtung und 2 in y-Richtung.

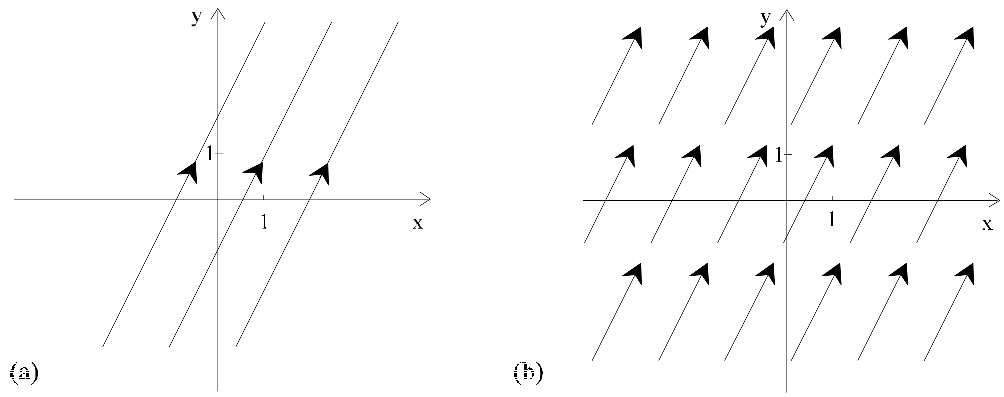

Abb. 1: (a) Trajektorien des Stromes $\varphi_t((x, y)) = (x + t, y + 2t)$; (b) zugeordnetes Vektorfeld.

Eine Trajektorie erhält man, indem man alle vergangenen und zukünftigen Werte eines Zustandes x_0 sammelt. Formal ist das die Menge $\{\varphi_t(x_0) \mid t \in \mathbb{R}\}$. Im Beispiel sind die Trajektorien alle Geraden im $\mathbb{R}^2$ mit der Steigung 2 (Abb. 1). Alle Trajektorien zusammen ergeben ein *Phasenportrait* oder *Phasenbild* des Systems. Will man ein Phasenportrait konkret zeichnen, so darf man natürlich nur einige ausgewählte Trajektorien(abschnitte) realisieren, da andernfalls das gesamte Diagramm schwarz einzufärben wäre.

An einer Stelle **x** hat ein Strom eine *Geschwindigkeit*, die sich durch

$$d\varphi_t/dt\,(\mathbf{x})\,|_{t=0} = \lim_{\varepsilon\to 0} (\varphi_\varepsilon(\mathbf{x}) - \varphi_0(\mathbf{x}))/\varepsilon = \lim_{\varepsilon\to 0} (\varphi_\varepsilon(\mathbf{x}) - \mathbf{x})/\varepsilon$$

ergibt. Dies ist ein Vektor, dessen Länge den Betrag, und dessen Richtung die Richtung der Bewegung von **x** im Strom kennzeichnet. Indem man diesen Geschwindigkeit bei jedem Punkt **x** des Zustandsraumes aufträgt, erhält man ein *Vektorfeld* **X**(**x**) (Beispiel in Abb. 1b).

C. Vektorfelder **X**(**x**) werden gern durch *Differentialgleichungen* spezifiziert. Dies ist in Anwendungen die häufigste Art, wie ein KDS dargestellt wird. Das System aus Abb. 1 etwa wäre durch die (sehr einfachen) Differentialgleichungen $dx/dt = 1$, $dy/dt = 2$ festgelegt. In Abb. 2 ist ein nicht ganz so triviales Beispiel gegeben.

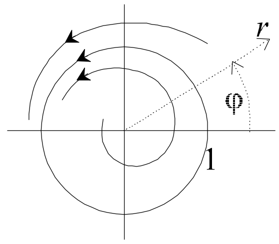

Abb. 2: Phasenportrait (in Polarkoordinaten) des dynamischen Systems $dr/dt = -r(r - 1)$, $d\theta/dt = 1$.

Systeme mit mehr als einer Dimension, z.B. zweidimensionale wie

$$dx_1/dt = g_1(x_1, x_2)$$
$$dx_2/dt = g_2(x_1, x_2)$$

werden oft durch eine einzige Vektorgleichung spezifiziert:

$$d\mathbf{x}/dt = \mathbf{g}(\mathbf{x}).$$

Zu jedem Strom gibt es ein äquivalentes System von DGL. Die Umkehrung gilt nicht: es gibt DGL, deren Lösungen nicht für alle Zeiten t angebbar sind (z.B. hat das eindimensionale System $dx/dt = x^2$ u.a. Lösungen $(C - t)^{-1}$, definiert nur für $-\infty \leq t \leq C$). Für manche theoretische Untersuchungen sind dennoch Ströme die Darstellung der Wahl.

Bemerkung zu A, B, C: Durch einen gegebenen Zustand x_0 verläuft stets genau eine Trajektorie: letztere "kreuzen" sich also nie. Bei Diffeomorphismen ist dies eine Folge der Bijektivität. Bei Strömen ist es eine Folge davon, daß die $\boldsymbol{\varphi}_t$ Diffeomorphismen sind. Für Differentialgleichungen $d\mathbf{x}/dt = \mathbf{g}(\mathbf{x})$ verlangt man gewöhnlich, daß $\mathbf{g}$ in allen Dimensionen des Phasenraumes partiell differenzierbar sei; dies garantiert die Kreuzungsfreiheit der Trajektorien.

Alle hier vorgestellten Darstellungen liefern *zeitinvariante* (synonym: *stationäre*, auch: *autonome* [was bei geeigneter Interpretation auch heißen kann: von Input unabhängige]) dynamische Systeme. Das bedeutet, daß sich die dynamische Gesetzmäßigkeit selbst (Diffeomorphismus, Strom, oder DGL) mit der Zeit nicht ändert. Solche dynamische Systeme können nicht altern, lernen oder sich sonstwie "substantiell" verändern. Es ist zwar formal möglich, DGL auch von der Zeit t selbst abhängig zu machen, indem t auf der rechten Seite als Parameter erlaubt wird. Solche nicht-stationären Systeme $d\mathbf{x}/dt = \mathbf{g}(\mathbf{x}, t)$. werden jedoch in der naturwissenschaftlichen Systemtheorie seltener und in der differentialgeometrischen Theorie dynamischer Systeme so gut wie nie untersucht. Eine zeitliche Änderung der Systemgesetzmäßigkeit selbst wird durch einen anderen Mechanismus berücksichtigt (Variation von Kontrollparametern, s. u.).

Eine oft verwendete Methode, die Untersuchung eines n-dimensionalen Stromes auf einen (n-1)-dimensionalen Diffeomorphismus hinunterzudrücken, sind *Poincaré-Schnitte.* Ein (globaler) Poincaré-Schnitt (engl: P. *section*) in einem n-dimensionalen Strom ist eine (n-1)-dimensionale Hyperebene H (allgemeiner: eine Untermannigfaltigkeit) mit den folgenden Eigenschaften:

1. Jede Trajektorie schneidet H unendlich oft,
2. keine Trajektorie berührt H tangential.

Intuitiv ist H ein Querschnitt durch den Strom, in dem die aufeinanderfolgenden "Treffer" wiederkehrender Trajektorien einen Diffeomorphismus **f** bestimmen. Präziser ist **f** definiert als die *erste Rückkehrabbildung*: Für $\mathbf{x} \in$ H ist $\mathbf{f}(\mathbf{x})$ definiert als der Punkt, in dem die Trajektorie durch **x** zum ersten Mal wieder nach H zurückkehrt. **f** heißt auch eine *Poincaré-Abbildung* (engl.: Poincaré map). In dem System aus Abb. 2 wäre jeder Radius ein globaler Poincaré-Schnitt mit der 1-dimensionalen Poincaré-Abbildung $\mathbf{f}(r) = r(r - 1)$. Nicht jedes dynamische System hat globale Poincaré-Schnitte (sogar die wenigsten). Man verwendet aber oft lokale Varianten.

Eines der zentralen Phänomene in dynamischen Systemen (nicht nur in kontinuierlichen) sind *Attraktoren*. Intuitiv ist ein Attraktor ein Gebiet im Phasenraum, das benachbarte Trajektorien bei $t \to \infty$ "anzieht", indem sie sich diesem Gebiet asymptotisch nähern. Abb. 3a zeigt den einfachsten Fall eines Attraktors, einen *Grenzpunkt* (limit point), und Abb. 3b den zweiteinfachsten Fall, einen *Grenzzyklus* (limit cycle). Man sieht, daß Trajektorien "aus der Umgebung" des Attraktors sich bei $t \to \infty$ beliebig nahe diesem nähern. In Abb. 3c laufen die Trajektorien parallel nebeneinander her, nichts nähert sich irgendetwas anderem - kein Attraktor in diesem Phasenportrait!

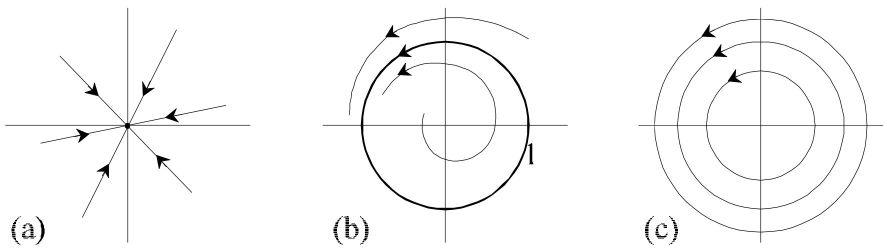

Abb. 3: Die beiden einfachsten Attraktor-Typen. (a) ein Grenzpunkt (Systemgleichung: $dx/dt = -x$, $dy/dt = -y$), (b) der Einheitskreis als Grenzzyklus (selbes System wie in Abb. 2), (c) *kein* Attraktor in diesem Phasenportrait! ($dx/dt = y$, $dy/dt = -x$)

Es gibt mit dem Grenzzyklus verwandte Erscheinungen, die etwas komplizierter aussehen als dieser. Sie entstehen, wenn sich mehrere Grenzzykel überlagern. Stehen die Periodenlängen der dabei beteiligten Grenzzykel nicht in einem rationalen Verhältnis zueinander, so ergibt sich keine zyklisch geschlossene Trajektorie mehr. Vielmehr liegt der dann *quasi-periodisch* genannte Attraktor dicht auf einem Torus in einem Phasenraum mit einer Dimension größer oder gleich drei (vgl. Haken 1983, 28f).

Die exakte Definition des Attraktorbegriffs ist ein bißchen verwickelt (Arrowsmith & Place 1992, 141). Da der Attraktorbegriff so fundamental ist, soll sie hier vorgestellt werden, auch wenn dies im weiteren nicht wichtig ist. Man definiert zunächst den Begriff der *trapping region* für einen Fluß $(\varphi_t)_{t \in \mathbb{R}}$ (im folgenden die englischen Begriffe, da mir keine deutschsprachigen Quellen für diese Definitionen bekannt sind). Eine *trapping region* ist eine kompakte, zusammenhängende Menge $\mathrm{T} \subseteq \mathbb{R}^n$, für die $\varphi_t(\mathrm{T}) \subseteq \mathrm{T}$ für alle $t \geq 0$.

Dann definiert man *attracting sets* A als solche abgeschlossene, invariante [d.h. $\varphi_t(\mathrm{A}) = \mathrm{A}$ für alle t] Gebiete des Phasenraumes, für die innerhalb jeder ε-Umgebung eine *trapping region* T existiert, so daß

$$\mathrm{A} = \bigcap_{t>0} \varphi_t(\mathrm{T}).$$

Schließlich definiert man als einen *Attraktor* ein *attracting set*, in welcher eine Trajektorie dicht liegt.

Die Komplexität dieser Definition ist der Existenz von "chaotischen" invarianten Mengen geschuldet, bei denen die Unterscheidung zwischen Attraktoren und Nicht-Attraktoren subtil sein kann.

Attraktoren sind in empirisch beobachteten dynamischen Systemen die herausragenden beobachtbaren Erscheinungen! Grob gesagt, sind "Regelmäßigkeiten" empirischen Systemverhaltens meistens als Attraktoren zu verstehen. Beispiele aus den Naturwissenschaften finden sich reichlich in Haken (1983). Beispiele aus dem Bereich der KI werden in Kapitel 4 vorgestellt.

Neben Punkt- und Grenzzykelattraktoren gibt es noch eine dritte Klasse. Sie wird erst seit den 60er Jahren systematisch untersucht, obwohl schon zu Beginn des Jahrhunderts verschiedentlich Mathematiker über dieses Phänomen "stolperten", sich aber, letztlich mangels Computergraphik, im wahrsten Sinne des Wortes kein Bild davon machen konnten. Es handelt sich um die *chaotischen* Attraktoren (zunehmend selten auch *seltsame* - engl. *strange* - Attraktoren genannt).

Zunächst ein Beispiel: der von Otto Rössler (1968) gefundene, nach ihm benannte, einfachste bekannte chaotische Attraktor. Seine Gleichung lautet $dx/dt = - y - z$, $dy/dt = x + ay$, $dz/dt = b + z(x - c)$. Für $a = .2$, $b = .2$, $c = 5.7$ ergibt sich ein dreidimensionaler Strom mit einem Rössler-Attraktor (Abb. 4).

Abb. 4: Eine Ausprägung des Rössler-Attraktors (aus Nicolis 1989, p. 333)

Chaotische Attraktoren sind Attraktoren mit einer fraktalen Struktur. Sie können erst ab Dimensionen des Zustandsraumesgrößer oder gleich 3 vorkommen. Ein zentrales Charakteristikum fraktaler Strukturen ist deren *fraktale Dimension*. Bei empirisch beobachteten Attraktoren bieten Nebenergebnisse der Abschätzungsprozedur für diese Größe einen Hinweis darauf, daß tatsächlich ein chaotischer Attraktor vorliegt (und nicht etwa bloßes Rauschen).

Außerdem liefert die fraktale Dimension Information darüber, wie kompliziert der Attraktor "tatsächlich" ist, d.h. wieviele Freiheitsgrade, grob gesprochen, das System effektiv ausnützt. Wenn man ein dynamisches System empirisch mißt, dann weiß man nicht von vorneherein, wieviele Dimensionen "eigentlich" das System hat — die "richtige" Anzahl von

Dimensionen muß keineswegs der Anzahl n der beobachteten Variablen entsprechen. Interessanterweise ist nun die fraktale Dimension, die man zu dem in n beobachteten Variablen beschriebenen Attraktor berechnen kann, unter gewissen (praktisch nicht einfach zu beherrschenden) Voraussetzungen unabhängig von n. Die Berechnung fraktaler Dimensionen kann so einen (unter Vorbehalt) von der zufälligen Beobachtungstechnik unabhängigen Einblick in das "Wesen" der zugrundeliegenden dynamischen Gesetzmäßigkeit liefern. Ein schönes Beispiel einer solchen Analyse findet man bei Robertson et al. (1993).

Hier soll nur der Begriff der fraktalen Dimension selbst erläutert werden. Sei $M \subset \mathbb{R}^n$ eine beschränkte Menge. Wir betrachten die Anzahl a der n-dimensionalen Würfel mit Kantenlänge r, die man braucht, um M zu überdecken. Wir definieren

$$D := \lim_{r \to 0} \ln(a) / \ln(1/r)$$

als die *fraktale Dimension* von M. Für "normale" m-dimensionale (wobei $m \leq n$) Mengen M sollte D gerade die Dimension m ergeben. Daß dies tatsächlich der Fall ist, zeigen die folgenden Beispiele.

Beispiel: $M = [0,1]\times[0,1]$ in $\mathbb{R}^2$, d.h. das Einheitsquadrat. Wir erwarten D = 2. Es ergibt sich tatsächlich (vgl. Abb. 5a):

$$\lim_{r \to 0} \ln(a) / \ln(1/r) = \lim_{r \to 0} \ln((1/r)^2) / \ln(1/r) = 2.$$

Beispiel: $M = [0,1]$ in $\mathbb{R}^2$, d.h. die Einheitsstrecke auf der x-Achse. Wir erwarten D = 1. Tatsächlich (vgl. Abb. 5b) ergibt sich:

$$\lim_{r \to 0} \ln(a) / \ln(1/r) = \lim_{r \to 0} \ln(1/r) / \ln(1/r) = 1.$$

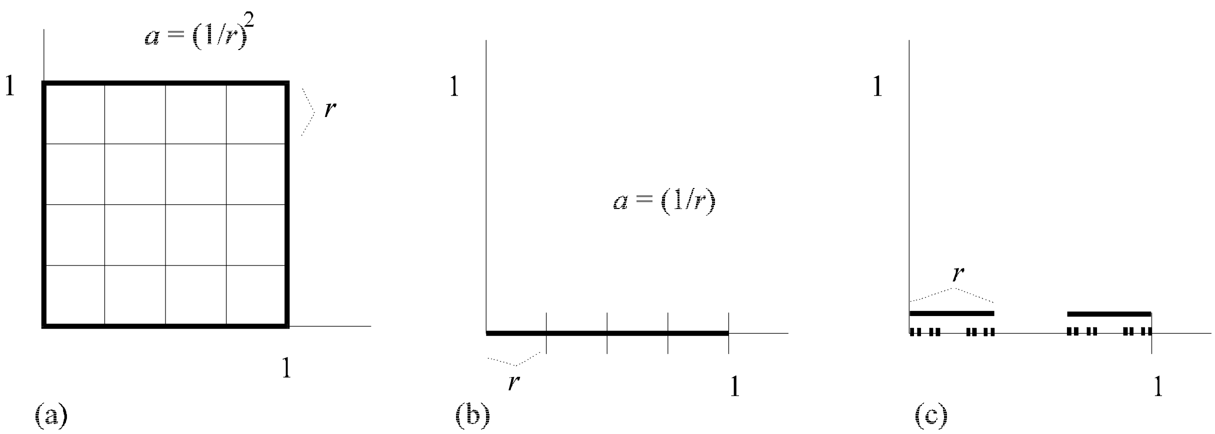

Abb. 5: Zur Berechnung der fraktalen Dimension des Einheitsquadrats (a) und der Einheitsstrecke (b) und der eindimensionalen Cantormenge (c) im $\mathbb{R}^2$.

Wenn wir aber die Cantor-Menge nehmen, die bekanntlich aus der Einheitsstrecke durch fortgesetztes Entfernen der mittleren Drittel der jeweils verbleibenden Strecken entsteht, erhalten wir folgende Tabelle (Abb. 5c zeigt $r = 1/3$):

r	a
1	1
1/3	2
1/9	4
...	...
$1/(3^n)$	2^n

Daraus ergibt sich dann

$$\lim_{r \to 0} \ln(a) / \ln(1/r) = \lim_{r \to 0} \ln(2^n) / \ln(3^n) = \lim_{r \to 0} \ln 2 / \ln 3 = .631...$$

Das ist aber keine ganze Zahl, die Cantor-Menge ist also ein Fraktal.

Diese Rechenvorschrift für D würde dieselben Ergebnisse liefern, wenn statt der n-Würfel etwa die n-dimensionale Kugel mit Radius r oder irgend eine andere Gestalt mit charakteristischer Kantenlänge r genommen würde.

Eine nicht-ganzzahlige fraktale Dimension ist eine hinreichende, aber keine notwendige Bedingung für das Vorliegen eines Fraktals. Es gibt auch Fraktale mit ganzzahliger fraktaler Definition (z.B. die Peano-Kurve, vgl. Kriz 1992 p. 56). Ferner gibt es eine Reihe von Varianten der oben gegebenen, klassischen Definition (einige davon in Robertson et al. 1993), sowie praktische Berechnungsmethoden, die nicht auf Überdeckungen beruhen.

Eine weitere Kenngröße für Attraktoren, die häufig zur Identifizierung chaotischer Attraktoren verwendet wird, sind *Lyapunov-Exponenten* (Haken 1983, 42ff, 79ff). Intuitiv entsprechen sie den exponentiellen Faktoren der Beschleunigung, mit der nahe beieinander liegende Trajektorien sich in einer gegebenen Richtung voneinander entfernen. Hier kann nur eine beispielhafte Beschreibung erfolgen.

Wir betrachten zunächst ein eindimensionales System $dx/dt = ax$. Die Lösungskurven dieses Systems sind im x-t-Koordinatensystem durch $x = Ce^{at}$ gegeben. Abb. 6 skizziert diese Kurven für negatives a, positives a, und den Fall $a = 0$. (Abb. 6 ist übrigens *kein* Phasenportrait; ein solches enthält keine t-Achse!)

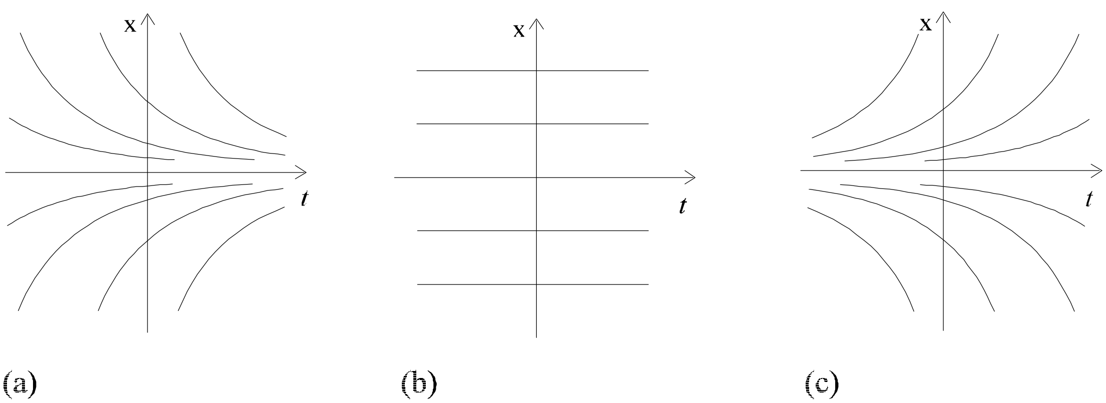

Abb. 6: Lösungskurven zu $dx/dt = ax$ für (a) $a < 0$, (b) $a = 0$, (c) $a > 0$.

In allen drei Fällen von Abb. 6 ist die Achse x = 0 eine Lösung. Der Wert x = 0 ist unabhängig von a ein Fixpunkt des Systems.

In (a) ist dieser Fixpunkt *labil* in dem Sinne, daß wenn zu einem beliebigen Zeitpunkt eine Trajektorie nahe bei x = 0 (aber nicht exakt im Wert x = 0) gestartet wird, sich diese exponentiell vom Fixpunkt entfernt. Anders ausgedrückt: kleine "Störungen" (Fluktuationen) führen zur Divergenz vom Fixpunkt. Der exponentielle Faktor dieser Divergenz ist dabei a.

In (b) ist der Fixpunkt *asymptotisch stabil* in dem Sinne, daß sich nahegelegene Trajektorien mit einem exponentiellen Faktor a (der hier negativ ist) auf den Fixpunkt zubewegen. "Störungen" werden wieder ausgeglichen.

In (c) schließlich ist der Fixpunkt *indifferent* (oder auch *stabil, aber nicht asymptotisch stabil*) — der exponentielle Änderungsfaktor benachbarter Trajektorien ist 0.

Die in diesem Beispiel auftretenden positiven, negativen, oder den Wert 0 annehmenden Exponenten *a* demonstrieren den einfachsten Fall von Lyapunov-Exponenten. Man sieht hier die Grundidee: LE messen den exponentiellen Divergenz- bzw. Konvergenzfaktor, mit der benachbarte Trajektorien sich von einer zu untersuchenden Lösung (hier: die Trajektorie x = 0) entfernen bzw. sich dieser nähern. Der LE gibt (in diesem Beispiel zunächst) darüber Auskunft, ob es sich bei dieser Trajektorie um einen Attraktor (LE kleiner 0), einen *Repellor* (LE größer 0), oder um weder noch handelt (LE = 0).

Für die allgemeine Definition von Lyapunov-Exponenten λ wird diese Grundidee verallgemeinert. Die einfache exponentielle Lösungsform $x = Ce^{at}$ ergibt sich nur bei *linearen* DGL mit konstanten Koeffizienten vom Typ $dx/dt = ax$. Um diese Situation auch bei beliebigen anderen DGL zu erreichen, benutzt man für die Definition von λ nicht die originale DGL, sondern nur die linearen Anteile davon (Jacobi-Matrix). Außerdem muß man berücksichtigen, daß sich das Divergenz/Konvergenzverhalten in der Nachbarschaft einer Trajektorie mit der Zeit ändern kann. Dies führt zu einer Supremumskonstruktion über die Zeit.

Zu einem Attraktor in $\mathbb{R}^n$ gibt es n Lyapunov-Exponenten (sie können auch numerisch zusammenfallen). Sie geben an, mit welchem exponentiellen Faktor sich nahebei liegende Trajektorien sich auf den Attraktor hinbewegen, was in n Dimensionen in n Richtungen verschieden schnell sein kann.

Wir betrachten zur Illustration einen Grenzzykel-Attraktor im $\mathbb{R}^2$. Hierzu nehmen wir wieder das Beispiel aus Abb. 2 (Gleichung $dr/dt = r(r - 1)$, $d\theta/dt = 1$ in Polarkoordinaten). Für den Grenzzykel auf dem Einheitskreis errechnet man zwei LE, nämlich $\lambda_1 = -1$ und $\lambda_2 = 0$ (λ_1 ist der Wert der Jacobi-Matrix von dr/dt an der Stelle $r = 1$, und λ_2 ist der Wert der Jacobi-Matrix von $d\theta/dt$, die konstant gleich 0 ist für alle θ). Der Wert $\lambda_1 = -1$ beschreibt, daß Trajektorien nahe des Einheitskreises sich diesem mit exponentiellem Faktor -1 nähern. Etwas komplizierter ausgedrückt: wenn x ein Punkt auf dem Einheitskreis ist, und y "quer" dazu, dann nähert sich die Trajektorie durch y der durch x mit exponentiellem Faktor -1. Das etwas rätselhaftere $\lambda_2 = 0$ beschreibt das Verhalten einer Trajektorie durch einen Punkt y, der selbst auf dem Einheitskreis liegt — die Trajektorien durch x und y halten auf Ewigkeit denselben Abstand, das führt analog zu Abb. 6c auf den Wert 0.

Man gibt die LE eines Attraktors üblicherweise der Größe nach geordnet an. Dabei interessieren nur die Vorzeichen: sie enthalten die wesentliche Information über die Natur der beschriebenen Trajektorie. Den Einheitskreis-Attraktor im eben untersuchten Beispiel würde man durch (-, 0) charakterisieren. (-, 0) charakterisiert in $\mathbb{R}^2$ stets zyklische Attraktoren. (-, -) würde Fixpunkt-Attraktoren in einem zweidimensionalen Zustandsraum charakterisieren.

Im in $\mathbb{R}^3$ charakterisiert (-, 0, 0) Torus-Attraktoren, (-, -, 0) eindimensionale Grenzzykel, und (-, 0, +) chaotische Attraktoren. Der positive LE bedeutet, daß in einer Richtung Divergenz vorliegt. Diese Divergenz muß irgendwie in den Attraktor "zurückgeführt" werden (sonst wäre es keine Attraktor). Bei der Abbildung des Rössler-Attraktors (Abb. 4) besteht dieses Zurückführen in der "hochgezogenen Kehrtkurve", welche Trajektorien aus der Peripherie der (in der Abbildung waagrecht angeordneten) divergenten, fraktalen Spirale in deren Zentrum zurückbringt.

Die (Wieder-)Entdeckung chaotischer Attraktoren, fraktaler Strukturen und damit verbundener Erscheinungen hat etwa ab Mitte der 60er Jahre zu einem ungeheuren Aufschwung in der Erforschung dynamischer Systeme geführt. Die landläufige, inzwischen zum geistigen Allgemeingut gehörende Bezeichnung "Chaostheorie" (die es als eigenständige Theorie gar nicht gibt) reflektiert diese Entwicklung (schöne semi-populäre Einführung in Phänomene des Chaos: Briggs & Peat 1989, Chaos in der Physik: Ford 1989). Weitreichende, wirklich neue und vor allem mathematisch fundierbare Einsichten in die Natur komplexer Systeme wurden im Laufe dieser Entwicklung gewonnen, z.B.:

- Physikalische Systeme, deren Verhalten auf den ersten Blick vollkommen erratisch erscheint, lassen sich oft als Systeme in chaotischen Attraktorzuständen deuten — sie weisen eine "verborgene Ordnung" auf und sind entgegen dem äußeren Anschein nur wenigen Freiheitsgraden bestimmt.
- Die Vorhersagbarkeit der Zukunft chaotischer Systeme ist durch die exponentielle Verstärkung kleiner Unbestimmtheiten der Anfangsbedingungen prinzipiell beschränkt (quantifizierbar durch Lyapunov-Exponenten). Dies ist der berühmte "Schmetterlings-effekt" mit seinen Folgen für die (Nicht-)Beherrschbarkeit komplexer (u.a. sozialer oder wirtschaftlicher) Systeme.
- Quer über äußerlich und auch mathematisch sehr unterschiedlich beschreibbare Systeme existieren universale Gesetzmäßigkeiten, die sich z.B. in bestimmten numerischen Konstanten niederschlagen (etwa der Feigenbaum-Konstante, s.u.).
- Chaotische Systeme sind zwar formal deterministisch, jedoch verhalten sie sich in puncto Vorhersagbarkeit in vielem wie stochastische Systeme. Dies führt zum Begriff des "deterministischen Chaos", der geeignet erscheint, einige alte erkenntnistheoretische Fragen neu zu interpretieren.

Heute sind viele Naturwissenschaftler schon generell geneigt, hinter jedem äußerlich als zufällig erscheinenden Prozeß eine verborgene chaotische Ordnung zu sehen. Eine Warnung davor, die Rolle von "echtem" Zufall geringzuschätzen, gibt Millonas (1994).

Nun sollen *Birfurkationen* vorgestellt werden. Zur Vorbereitung benötigen wir die beiden Begriffe der *topologischen Äquivalenz* und der *Kontrollparameter*.

Zwei dynamische Systeme S_1 und S_2 heißen topologisch äquivalent, wenn das Phasen-portrait von S_1 durch eine stetige Bijektion so auf das Phasenportrait von S_2 abgebildet werden kann, daß Trajektorien auf Trajektorien abgebildet werden, wobei deren Orientierung erhalten bleibt. Man kann sich das intuitiv so vorstellen, daß das Phasenportrait von S_1 auf ein Blatt aus Gummi aufgemalt wird, und dieses dann verzerrt wird, um das Phasenportrait von S_2 zu erhalten. So sind etwa die beiden Phasenportraits in Abb. 7a topolgisch äquivalent.

Was den Begriff der Kontrollparameter betrifft, so erinnern wir uns, daß die DGL (oder Diffeomorphismen), mit denen dynamische Systeme spezifiziert werden, üblicherweise — neben den Zustandsvariablen — noch eine weitere Sorte Parameter enthalten, die hier bisher stillschweigend als Konstante aufgefaßt wurden. So ist zum Beispiel in $dx/dt = ax$ (vgl. Abb. 6) x eine Zustandsvariable, und a eine "Konstante". Wenn man nun Phasenportraits "verzerren" möchte wie eben angedeutet, dann kann man dies technisch durch eine Variation dieser "Konstanten" bewerkstelligen. Man untersucht einfach verschiedene "Ausführungen" des dyna-mischen Systems, die sich durch verschiedene Werte dieser Konstanten ergeben. In mathematischer Ausdrucksweise: man untersuch eine parametrisierte Familie von dynamischen Systemen. Die zu variierenden "Konstanten" nennt man in diesem Zusammenhang Kontroll-parameter.

Ein dynamisches System kann die Eigenschaft haben, *strukturell stabil* zu sein. Das bedeutet intuitiv, daß kleine Störungen seinen topologischen Charakter nicht verändern. Präziser nennt man ein dynamisches System strukturell stabil, wenn man sein Vektorfeld mit beliebigen, aber hinreichend kleinen (im Sinne einer ε-Umgebung) Vektorfeldern überlagern kann, so daß die resultierenden Systeme topologisch äquivalent zum Ausgangssystem sind. Abb. 7b exemplifiziert dies Prinzip.

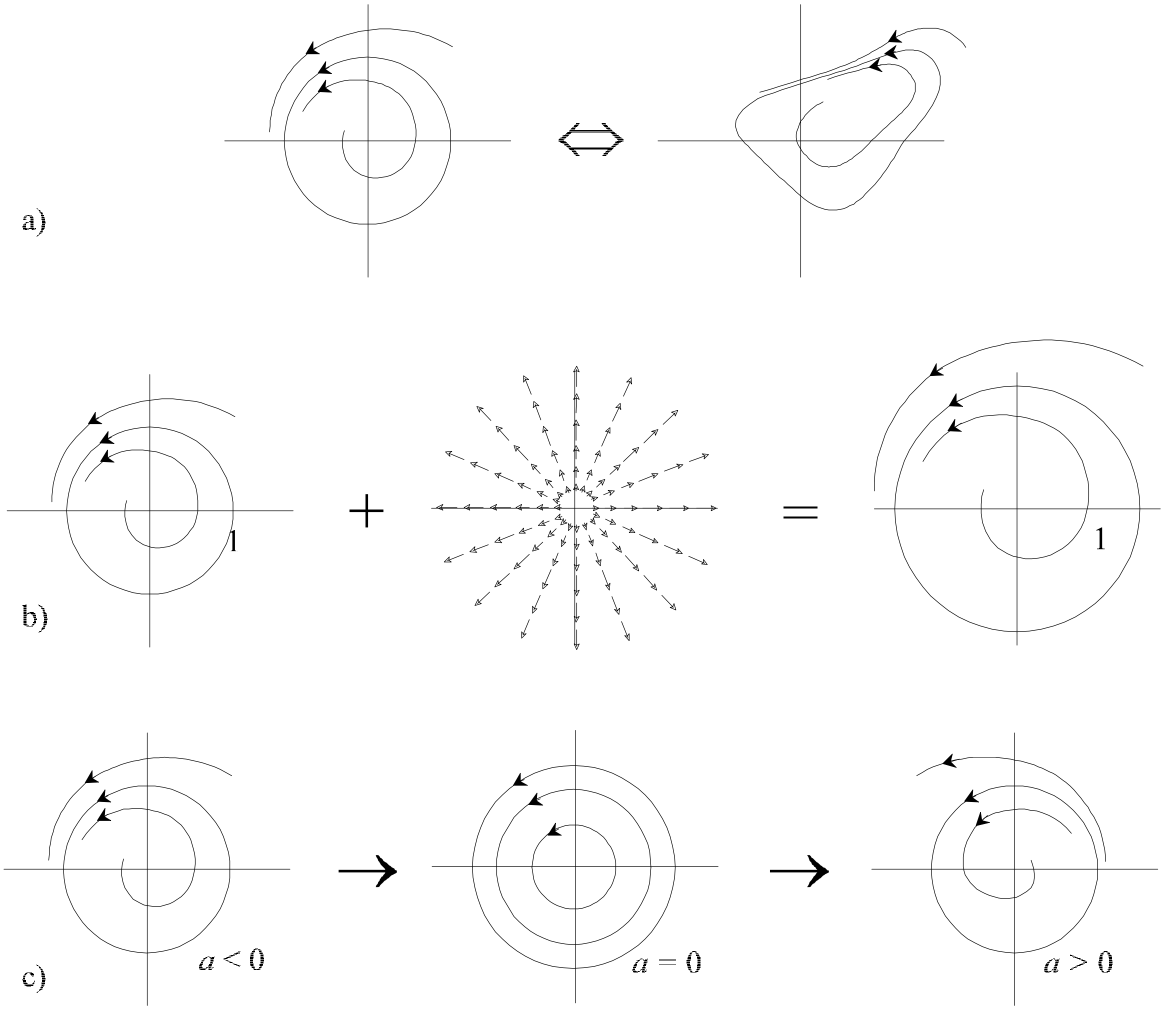

Abb.7: Topologische Äquivalenz (a), strukturelle Stabilität (b), und eine Bifurkation (c).

Strukturelle Stabilität ist eine weit verbreitete Eigenschaft unter dynamischen Systemen (mehr zu diesem sehr wichtigen Gegenstand der mathematischen Systemtheorie bei Abraham & Shaw 1983, Kap. 11 und 12, und bei Arrowsmith & Place 1990, Kap. 3). Es gibt jedoch auch qualitativ instabile Systeme, bei denen beliebig kleine Veränderungen des Vektorfeldes zu einem qualitativen Umschlag des Phasenportraits führen können. Solche instabilen Systeme markieren sog. *Bifurkationspunkte* (genaue Definitionen bei Arrowsmith & Place 1990, Kap. 4). Technisch wird die hier angesprochene kleine Verformung von Vektorfeldern zumeist durch die ε-kleine Variation von Kontrollparametern bewerkstelligt.

Typischerweise erhält man folgendes Gesamtbild: Bei Variation der Kontrollparameter von einen Ausgangswert her ändert sich das qualitative Bild des dynamischen Systems zunächst nicht (entsprechend seiner strukturellen Stabilität), obwohl es zu quantitativen Verschiebungen kommt (wie in Abb. 7b). Doch wenn die Kontrollparameter bestimmte kritische Werte überschreiten, kommt es zu einem plötzlichen qualitativen Umschwung, einer "Neuorganisation" des Systems, die dann wieder eine Weile strukturell stabil bleibt. Diese schlagartigen Reorganisation nennt man *Bifurkation*. Abb. 7c zeigt einen solchen Übergang zwischen zwei strukturell stabilen Systementypen, die über einen strukturell instabilen Bifurkationspunkt führt. Die Systemgleichung lautet in diesem Beispiel (ähnlich wie in Abb. 2)

$$dr/dt = ar(r - 1)$$
$$d\theta/dt = 1$$

Für $a < 0$ erhält man Phasenportraits wie das linke in Abb. 7c, d.h. mit einem zyklischen Attraktor und einem Punkt-Repellor im Ursprung. Für alle negativen a sind die erhaltenen Systeme topologisch äquivalent. Für alle $a > 0$ erhält man entsprechend Phasenportraits mit einem zyklischen Repellor und einem Punkt-Attraktor. Für den singulären Wert $a = 0$ erhält man eine Familie von zyklischen Trajektorien um einen Fixpunkt im Ursprung. Diese Trajektorien sind keine Attraktoren (vgl. Abb. 3). Der Wert $a = 0$ markiert einen Bifurkationspunkt, denn in jeder ε-Umgebung von 0 liegen offenbar Werte von a, für die das zugehörige System nicht topologisch äquivalent zu dem System mit $a = 0$ ist.

Bei wiederholtem Vor- und Zurückfahren des Kontrollparameters über einen kritischen Wert kommt es oft zu *Hysterese*-Effekten, d.h. der Umschlag in einer Richtung erfolgt bei einem anderen Wert als der Umschlag in der Rückrichtung (einfaches Beispiel in Kelso et al. 1993).

Es gibt viele gut untersuchte Typen von Bifurkationen. Hier sollen zwei davon beispielhaft skizziert werden: Periodenverdopplungen und Hopf-Bifurkationen.

Abb. 8: Trajektorien des Verhulst-Prozesses für verschiedene r (aus Peitgen & Richter 1986, p. 24).

Von *Periodenverdopplungen* spricht man, wenn bei Variation von Kontrollparametern ein Grenzzykel-Attraktor A in einen anderen Grenzzykel-Attraktor umspringt, und zwar so, daß gleichermaßen aus einem einzigen "Umlauf" der Trajektorie in A zwei verschiedene, alternerierende Umläufe werden, die zusammen B ergeben. Für ein klassisches diskretes Beispiel sei der eindimensionale Diffeomorphismus

$$\mathbf{f}(\mathrm{x}) = (1 + r)\mathrm{x} - r\mathrm{x}^2, \text{ bzw.}$$
$$\mathrm{x}_{t+1} = (1 + r)\mathrm{x}_t - r\mathrm{x}_t^2$$

betrachtet (nach Peitgen & Richter 1986, Kap. 1; ein einfacheres, aber nicht so klassisches Beispiel ausführlich in Haken 1983, 50ff). Dies ist eine u.a. aus der Populationsdynamik bekannte Gleichung, welche dort eine Populationsgröße x_{t+1} auf die Populationsgröße x_t (etwa im Vorjahr) zurückführt. Der Term $(1 + r)\mathrm{x}_t$ entspricht exponentiellem Wachstum, und der Term $-r\mathrm{x}_t^2$ modelliert die "Bremsung" dieses Wachstum aufgrund von Ressourcenknappheit. Dies System wird auch "Verhulst-Prozeß" genannt.

In dieser Interpretation der Gleichung ist *r* die Reproduktionsrate. Sie ist formal der Kontrollparameter. Wenn man den Prozeß mit $x_0 = .1$ startet und verschieden *r* betrachtet, erhält man die Trajektorien aus Abb. 8.

Für *r* = 1.8 pendelt sich die Populationsdichte auf einen stabilen Wert ein. Für *r* = 2.3 findet man von Jahr zu Jahr stabil alternierende Werte. Diese Periode "verdoppelt" sich beim Übergang zu *r* = 2.5. Für *r* = 3 schließlich ist keine Regelmäßigkeit mehr erkennbar.

Wenn man die Werte für *r* auf der Abszisse, und die nach einer "Einschwingzeit" gefundenen Werte der x_t auf der Ordinate abträgt, erhält man das sog. *Feigenbaum-Diagramm* (Abb. 9).

Abb. 9: Das Feigenbaum-Diagramm (Aus Kriz 1992, p. 35).

Wichtige Eigenschaften des Feigenbaum-Diagramms sind im folgenden gesammelt:

- Für kleine *r* ist nur ein x-Wert stabil: es liegt ein einfacher Fixpunktattraktor vor.
- Mit wachsemdem *r* kommt es bei *r* = 2 zu einer plötzlichen qualitativen Verhaltensänderung, einer Bifurkation: ein einfach periodischer Attraktor entsteht.
- Bei weiter wachsendem *r* kommt es zu weiteren Periodenverdopplungs-Bifurkationen, deren Abstände kürzer werden. Tatsächlich ist im Limes jede folgende qualitativ stabile Region um den Faktor ≈ 4.669 (*Feigenbaum-Konstante*) kürzer als die vorherige.
- Bei *r* ≈ 2.570 gibt es eine Bifurkation anderen Typs: das Verhalten wird auf einen Schlag aperiodisch. Es erscheint völlig irregulär; kleinste Änderungen von Startwerten führen auf völlig verschiedene Trajektorien. Der Prozeß ist *chaotisch* geworden.
- Auch im Bereich jenseits von *r* = 2.570 gibt es weitere Bifurkationen. Deren Natur ist jedoch noch nicht vollständig aufgeklärt (vgl. Abraham & Shaw 1983f, Band IV, p. 25-46, für eine detailliertere, anschauliche Beschreibung).

Feigenbaum-Diagramm und Feigenbaum-Konstante entstehen nicht nur aus der Gleichung, welche die Verhulst-Dynamik beschreibt, sondern auch bei vielen anderen. Zuhause beobachtbar ist z.B. eine Periodenverdopplung beim Tropfrhythmus eines Wasserhahns (Kontrollparameter: Regulierung des Ventils; funktioniert nur bei fein regulierbaren Exemplaren ohne Turbulator). Geschichtlich ist hier ist zum ersten Mal eine *universelle* Gesetzmäßigkeit gefunden worden in einem Bereich, den man vorher nicht bzw. als stochastisches Rauschen mißverstand. Der steile Aufstieg und die große Faszination der "Chaostheorie" (die es als eigenständige wissenschaftliche Disziplin, wie gesagt, eigentlich gar nicht gibt) nahmen u.a. von hier ihren Ausgang, und das Feigenbaum-Diagramm ist zur Ikone geworden.

Periodenverdopplungen gibt es auch in einer kontinuierlichen Variante in Strömen (vgl. Haken 1983, 39f), und es gibt auch Periodenverdreifachungen, gemischte Formen, etc. Insgesamt haben wir es hier mit einem besonderen "Weg ins Chaos" (engl. *route into chaos*) zu tun. Speziell die Synergetik beschäftigt sich mit dieser Frage: wie gelangt ein System, bei

Variation von Kontrollparametern, von einem "geordneten" in einen "chaotischen" Zustand? Außer dem Periodenverdoppelungs-Szenario sind inzwischen eine Reihe z.T. gänzlich andersartiger solcher Wege ins Chaos bekannt — und auch verschiedene Sorten Chaos (Haken 1983, p. 264ff).

Hopf-Bifurkationen liegen vor, wenn ein Fixpunkt-Attraktor in einen Grenzzyklus-Attraktor umschlägt. Solche Bifurkationen sind in Anwendungen häufig. Das klassische Beispiel ist die Anregung eines elektrischen Schwingkreises. Hopf-Bifurkationen werden in einer ausgedehnten Literatur untersucht (anschauliche Einführung: Abraham & Shaw 1983f, Band IV, mathematisch-praktische Einführung: Haken 1983, 230ff, mathematisch abstrakte Behandlung: Arrowsmith & Place 1990, 234ff). Es gibt zwei Varianten: die Schwingung kann "sanft" mit zunächst beliebig kleiner Amplitude einsetzen (*soft excitation*, superkritische Hopf-Bifurkation), oder sofort mit einer Amplitude einer gewissen Größe (*hard excitation*, subkritische Hopf-Bifurkation).

Hier sei ein formales Beispiel für den ersten Fall vorgestellt. Sei (in Polarkoordinaten) $dr/dt = r(a - r^2)$, $d\theta/dt = 1$. Wie man aus der Funktion $r(a - r^2)$ sieht (vgl. Abb. 10a), ist für $a \leq 0$ der Ursprung ein Fixpunktattraktor (Phasenportrait in Abb. 10b). Für $a \geq 0$ jedoch erhält man einen Grenzzykel-Attraktor, nämlich einen Kreis um den Ursprung mit Radius $a^{1/2}$ (Abb. 10c). Diese Bifurkation ist eine "soft excitation", weil der Radius des Grenzzyklus für $a \downarrow 0$ selbst gegen 0 geht.

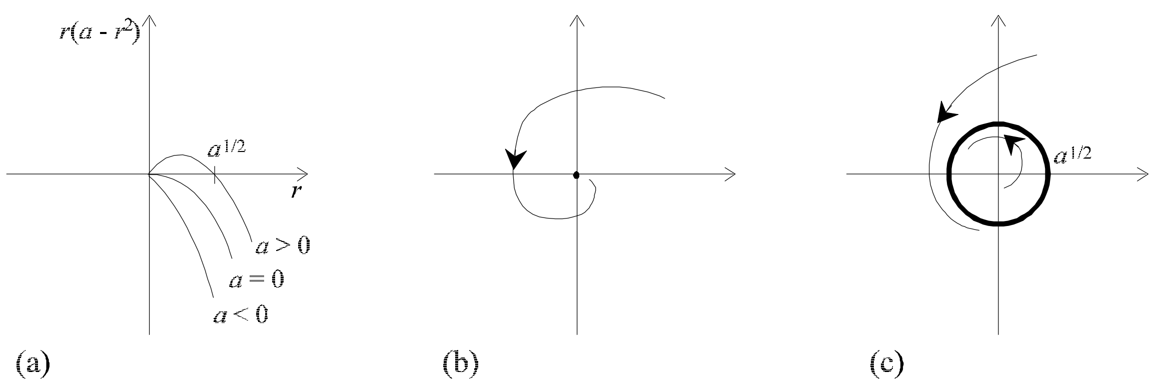

Abb. 10: Eine superkritische Hopf-Bifurkation: das System $dr/dt = r(a - r^2)$, $d\theta/dt = 1$.

Zum Abschluß dieses Kapitels soll das sog. *Versklavungsprinzip* vorgestellt werden. Es ist ein originärer Beitrag der Synergetik zum Verständnis komplexer Systeme (Haken 1983, p. 35f, Kapitel 7). Kurz gefaßt, besagt es, daß gewisse (wenige) Systemvariable die restlichen (vielen) dominieren, in dem Sinne, daß sich aus der Kenntnis der Dynamik der ersteren die Dynamik der letzteren ableiten läßt. Die wenigen, dominierenden Variablen werden auch *Ordnungsparameter* genannt (nicht zu verwechseln mit Kontrollparametern), und man sagt, sie *versklaven* die restlichen Parameter. Eine Vorbedingung für eine solche Versklavung ist, daß die Ordnungsparameter sich in einem wenig stabilen Bereich in der Nähe einer Bifurkation befinden.

Das Standardbeispiel ist der Laser. Wenn er "lasert", schwingen intuitiv gesprochen alle seine atomaren Mikro-Oszillatoren im Gleichtakt und in einer Richtung - zur Beschreibung des Zustandes reicht also eine einzige Variable aus. Befindet sich der Laser dagegen im unterkritischen Zustand, so schwingen alle Mikro-Oszillatoren im wesentlichen unabhängig voneinander und müssen alle einzeln beschrieben werden.

Hier soll nur an einem einfachen Beispiel demonstriert werden, wie nahe einer Bifurkation eine instabile (*unstable*) Variable u eine stabile Variable s versklavt. Dazu werde das folgende System betrachtet (vgl. Haken 1983, p. 187):

$$du/dt = au - us$$
$$ds/dt = -bs + u^2$$

Man beginnt mit der Annahme, daß sich u mit der Zeit kaum verändert. Ein solches Verhalten ist charakteristisch für die Nähe einer Bifurkation (*critical slowdown*): der dominierende lineare Teil au in $au - us$ wird klein für kleines a, was aber gerade die Nähe zu einer Bifurkation (die bei $a = 0$ vorliegt) bedeutet.

Im Extremfall bleibt u näherungsweise konstant. Für konstantes u sieht die Dynamik von s aus wie in Abb. 11 angegeben.

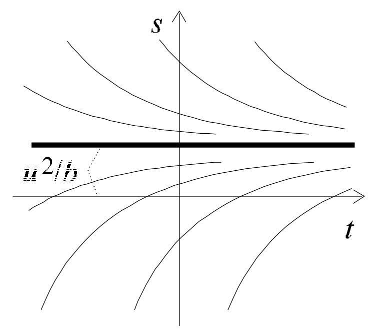

Abb. 21: Die Dynamik von s bei konstantem u.

Die Geschwindigkeit der Dynamik von s wird im wesentlichen durch den Faktor $-b$ in ds/dt-= $-bs + u^2$ bestimmt. Wenn sich u viel langsamer verändert, als es dieser Dämpfungsgeschwindigkeit von s entspricht, dann ist u in bezug auf s "im wesentlichen" konstant, und u zieht s nach sich entsprechend der Beziehung von Abb. 11, d.h., die Dynamik von s ergibt sich aus der von u gemäß

$$s(t) = u^2(t) / b.$$

Das Gesamtbild kann folgendermaßen beschrieben werden: Wenn Kontrollparameter (hier: a) gewisse Systemvariable u in die Nähe einer Bifurkation bringen, kann der Fall eintreten, daß diese u zu Ordungsparametern werden und die restlichen Systemvariable s versklaven.

4 Dynamische Systeme in der Modellierung intelligenter Informationsverarbeitung

Dies Kapitel bringt einen ansatzweisen Überblick über den Stand der Kunst bei der Verwendung dynamischer Systeme zur Modellierung intelligenter Informationsverarbeitung. Da systemtheoretische Methoden in der KI selbst noch kaum rezipiert sind, werden vor allem Arbeiten aus den Nachbarwissenschaften herangezogen: Kognitionswissenschaft und Psychologie, Linguistik, Robotik, Neurowissenschaften, Neuronale-Netz-Forschung.

Das Kapitel ist in vier Teile gegliedert. Zuerst wird die gegenwärtige Bedeutung systemtheoretischer Methoden in den einzelnen Disziplinen knapp umrissen (Abschnitt 4.1). Den Hauptteil (Abschnitt 4.2) nimmt die Darstellung einzelner inhaltlicher Aspekte intelligenter Informationsverarbeitung ein. Es wird referiert, wie systemtheoretische Methoden, quer über die Fachdisziplinen, intelligenzbezogene Phänomene erschließen. In zwei ergänzenden Abschnitten finden sich Bemerkungen über die Beziehung zwischen symbolischen und systemtheoretischen Modellen (4.3) sowie über die (nicht-)reduktionistische Natur systemtheoretischer Erklärungen.

4.1 Systemtheoretische Methoden in den einzelnen Fachdisziplinen: eine knappe Orientierung

In der **Kognitionswissenschaft** und der **Psychologie** spielen systemtheoretische Methoden bei der Erklärung motorischen Verhaltens seit Beginn der 80er Jahre eine zunehmend wichtige Rolle. Der Impetus ist dabei wohl ursprünglich von der Synergetik in die Psychologie hineingetragen worden (cf. Referenzen in Schöner et al. 1986). Die formale Modellierung motorischer Systeme als dynamische Systeme ist inzwischen etabliert, und die hier erzielten Erkenntnisse inspirieren die Übertragung der rezipierten Methoden in andere Teilgebiete der Kognitionswissenschaft und Psychologie (Sammelband: Smith & Thelen 1993a). Neuerdings werden dynamische Systeme sogar als ein fundamentales Paradigma zur Modellierung kognitiver Prozesse vorgeschlagen (van Gelder & Port 1994b). Ein besonders aktives Zentrum dieser Bewegung ist die University of Indiana in Bloomington. Neben diesen eigenständig der Kognitionswissenschaft und der Psychologie zuzurechnenden Entwicklungen werden formale Methoden dynamischer Systeme durch konnektionistische Techniken in das Gebiet importiert.

In der **Linguistik** gibt es traditionell Denkrichtungen mit einer gewissen Geistesverwandtschaft zu systemtheoretischen Methoden. Zu nennen ist hier zuvörderst der Einfluß des Strukturalismus: *"That language is a system is undoubtedly agreed upon by the vast majority of linguists today.* [...] *Clearly, the characterization stems from structuralism and can be used, in the first place, to emphasise a programmatic approach: that the role of the elements should merely be analysed as being dependent on the relations among them — regardless of their form. [...] the structuralistic approach to language as a system had necessarily to be a static one."* (Köhler 1987, 241f). Auf den Kern reduziert, sieht die strukturalistische Tradition die Sprache zwar als System, aber nicht als dynamisches System. Hierzu gibt es eine Gegenbewegung, die den dynamischen Charakter sprachlicher Prozesse betont (Sammlung in Rieger 1985), ohne allerdings eigentliche systemtheoretische Methoden einzuführen. Es gibt jedoch vereinzelte Versuche, dies zu tun. Köhler (1987) schlägt vor, Sprachen als dynamische Systeme im Sinne der Synergetik aufzufassen. Rickheit & Strohner (1992) entwickeln ein systemtheoretische Sicht nicht nur auf Sprachen als abstrakte Forschungsobjekte der klassischen Linguistik, sondern auch auf kognitive und kommunikative Prozesse als Gegenstände der Psycholinguistik. Tucker & Hirsh-Pasek (1993) beschäftigen sich mit der Frage des Spracherwerbs bei Kindern. Sie geben eine breite Übersicht über bestehende

Theoriebildungen und empirische Befunde, und zeigen, daß eine systemtheoretische Interpretation (im Sinne der Synergetik) einige schwierige Fragen beantworten kann. Bei Köhler, Rickheit & Strohner, und Tucker & Hirsh-Pasek geht es um die Etablierung eines methodischen Rahmens und um Begriffsklärungen; eine mathematische Formalisierung findet noch nicht statt. Den methodisch und mathematisch am weitesten entwickelten Ansatz liefert Wildgen (1985). In engem Bezug auf die Katastrophentheorie von René Thom bildet er einige Grundtypen von Bifurkationen sowohl auf raumzeitliche Prozesse in der physikalischen Außenwelt, als auch auf grammatische Strukturen ab und stellt so eine Verbindung zwischen Sprache und Welt dar. Strengere systemtheoretische Techniken gelangen ab und zu als "Irrgast" durch konnektionistische Methoden in die Linguistik (z.B. Smolensky 1986, Legendre et al. 1990a,b), bleiben aber für das methodologische Selbstverständnis der Disziplin vorerst ohne Folgen.

In der **Neurophysik**, **Biokybernetik** und der **Neuronale-Netze-Forschung** bilden Methoden der Kybernetik, Kontrolltheorie, der naturwissenschaftlichen und der mathematischen Theorie dynamischer Systeme die Basis der meisten formalen Modelle. Munro & Anderson (1988) schlagen sogar vor, die Ausbildung in konnektionistischen Techniken auf der formalen (Simulations-)Äquivalenz neuronaler Netze mit dynamischen Systemen aufzubauen. Viele einschlägige Arbeiten werden von Physikern (mit-)geschrieben. Das mathematische Niveau ist entsprechend hoch. Es gibt mehrere einschlägige Zeitschriften und eine Vielzahl spezialisierter Forschungspfade. Eine Übersicht kann hier nicht versucht werden. Arbeiten aus dem Gebiet werden im folgenden Abschnitt 4.2 reichlich zitiert.

In der **Robotik** gehören Methoden der Kontrolltheorie naturgemäß zum Standardrepertoire ingenieursmäßiger Techniken der Bewegungskontrolle und der Signalverarbeitung (z.B. in reiner Form bei Triggs 1994, in Kombination mit neuronalen Netzwerken bei Schweitzer & Wen 1994). Solche Methoden sind modellbasiert, d.h. sie setzen eine explizite mathematische Modellierung des Bewegungsapparates, der Sensorcharakteristiken etc. voraus. In den letzten Jahren sind auch Prinzipien der Selbstorganisation ausgenutzt worden, oft in Anlehnung an biologische Vorbilder (z.B. Gaudiano & Grossberg 1991, Cruse et al. 1994).

Die Erforschung **autonomer adaptiver Agenten** entwickelt sich zu einem eigenständigen Gebiet zwischen Biologie, Kognitionswissenschaft, KI und Robotik. Systemtheoretische Ansätze kommen hier in mindestens dreifacher Weise zum Tragen. Erstens sind jüngsthin Vorschläge gemacht worden, Agent-Umgebungs-Systeme formal durch gekoppelte, kontinuierliche dynamische Systeme zu modellieren (Smithers 1994, Beer 1995). Vor diesem Hintergrund werden einzelne motorische Behaviors als Attraktoren in dynamischen Systemen analysiert (Steels 1994b, Beer 1995). Diese Formalisierungen können als vorläufige Verwirklichung von programmatischen Äußerungen gelten, die in der epistemologischen Debatte um das *Situated Action*-Paradigma von dessen Proponenten gemacht worden sind (z.B. Clancey 1993, Greeno & Moore 1993). Zweitens werden Techniken der Kontrolltheorie durch die Robotik und neuronale Netze in das Gebiet importiert, wenn es um den Bau physikalischer Roboter oder die Analyse der mechanischen Leistungen biologischer Agenten geht (z.B. Weidemann & Pfeiffer 1994, Steinkühler & Cruse 1994). Drittens werden durch die Thematik Adaptivität/Lernen in Einzelfällen systemtheoretische Gedanken importiert (z.B. evolutionäre Techniken bei Steels 1994c, diskrete dynamische Systeme als kontrolltheoretisches Modell der zu lernenden Umgebung bei Kaelbling 1991). Insgesamt können zumindest dem programmatischen Anspruch nach systemtheoretische Methoden als etabliert gelten, wenn auch die konkrete Ausgestaltung gerade erst begonnen hat.

In der klassischen **künstlichen Intelligenz**, die sich mit "höheren", rationalen und begrifflichen Intelligenzphänomenen auseinandersetzt, sind systemtheoretische Methoden noch so gut wie gar nicht in Erscheinung getreten. Mir sind nur zwei Fälle bekannt. Hasida (1994) implementiert eine Kontrollstrategie für fuzzy Prolog-Programme, welche auf einem Energieminimierungsprinzip in einem kontinuierlichen dynamischen System beruht. Und ich selber (Jaeger 1994a,b,c,d) entwickele einen Typ diskreter, selbstorganisierender sog. *dynamischer Symbolsysteme* als formales Modell intelligenter Agenten, die von der sensomotorischen Peripherie bis zu einer begrifflichen Verarbeitungsebene einheitlich beschrieben werden können. Eine Sonderrolle nehmen die Arbeiten von Sacks (1990, 1991) ein. In einem *qualitative reasoning*-Kontext beschreibt er, wie mit klassischen KI-Schlußfolgerungsmethoden *über* planare kontinuierliche Systeme räsoniert wird.

4.2 Systemtheoretische Beiträge zum Verständnis intelligenter Systeme

Auf allen Beschreibungsebenen haben systemtheoretische Methoden bereits Einsichten in die Natur intelligenter Systeme geliefert. Ohne Anspruch auf Vollständigkeit wird in diesem Abschnitt eine Auswahl interessanter Arbeiten ausgebreitet, die einen Eindruck von den spezifischen Stärken und der Vielseitigkeit systemtheoretischer Methoden vermitteln soll. Die Darstellung beginnt bei physik-nahen, neuronalen Phänomenen (4.2.1), und vollzieht dann einen Aufstieg über sensomotorische Phänomene (4.2.2) und die konzeptuelle Verarbeitungsebene (4.2.3) bis zur Ebene von kompletten Agent-Umwelt-Systemen (4.2.4). Zur Vervollständigung wird noch auf die Beziehungen zwischen *computation* und physikalischer Dynamik hingewiesen (4.2.5).

Die nun zu referierenden Arbeiten stammen aus verschiedenen Fachdisziplinen. Um dem Leser die Zuordnung zu ermöglichen, wird der disziplinäre Hintergrund der Autoren jeweils durch Kürzel mit angegeben (subjektive Einschätzung — ich bitte die Betroffenen um Verzeihung). Es bedeuten: *AA* = autonome Agenten, *Bio* = Biokybernetik und Neurowissenschaft, *Inf* = Informatik, *KI* = Künstliche Intelligenz, *Ling* = Linguistik, *Kog* = Kognitionswissenschaft, *Math* = Mathematik, *NN* = Neuronale-Netz-Forschung, *Phil* = Philosophie, *Phys* = Physik, *Psych* = Psychologie, *Rob* = Robotik und ingenieursmäßige Mustererkennung, *Soz* = Sozial- und Wirtschaftswissenschaft.

4.2.1 Neuronale Prozesse

Wenn man irgendein Objekt als dynamisches System analysieren möchte, so besteht der erste Schritt in der Aufstellung eines plausiblen mathematischen Modelles, z.B. eines Systems von Differentialgleichungen.

Bei künstlichen neuronalen Netzen ist dieser erste Schritt im Vergleich zu anderen Untersuchungsobjekten in den meisten Fällen außergewöhnlich einfach. Als Zustandsvariable bieten sich kanonisch die Aktivierungen der einzelnen Neurone an, die auch fast immer gewählt werden. Die Spezifikation der Dynamik ergibt sich ohne weiteres aus den Aktivierungsfunktionen der Neurone.Diese "Modellierung" und ihre Varianten ist so direkt, daß künstliche neuronale Netze als eine notationelle Spielart einer großen Klasse von dynamischen Systemen gesehen werden können. Entsprechend schlagen Munro & Anderson (*NN* 1988) auch vor, die akademische Lehre im Bereich neuronale Netze durchgängig auf der Interpretation von neuronalen Netzen als dynamischen Systemen aufzubauen. Man kann vermuten, daß der große Erfolg der Neuronale-Netz-Forschung unter anderem auf dieser Auffaßbarkeit von neuronalen Netzen als dynamische Systeme beruht. Sie erlaubt es, bekannte, starke systemtheoretische Methoden ohne die Probleme von mehr oder weniger willkürlichen Modellierungsentscheidungen anzuwenden.

Ob man ein künstliches neuronales Netz in seiner originalen Spezifikation oder in einer kanonischen Reformulierung z.B. durch Differentialgleichungen untersucht, ist also weitgehend

Geschmackssache. In beiden Fällen lassen sich sowohl formale mathematische Techniken als auch Computersimulationen zur weiteren Analyse verwenden.

Bei natürlichen neuronalen Netzwerken ist die Lage diffiziler. Da viele strukturelle und physiologische Details biologischer Netzwerke unbekannt sind, ist man auf vereinfachende Modellbildungen angewiesen. Eine zentrale Vereinfachung ist, daß die Funktion eines Neurons in einem komplexen Netz durch einen einzigen Aktivierungswert fixiert werden kann. Diese Vereinfachung blendet möglicherweise wesentliche Aspekte biologischer Neurone aus ("realistische" Modellierung mit beeindruckend reicher funktionaler Phänomenologie in De Schutter & Bower *Bio* 1994), sie ist aber gängige Praxis (mit Ausnahmen, vgl. kurze Diskussion in Gaudiano & Grossberg *Bio*, *NN* 1991 p. 180). Nach dieser Vor-Vereinfachung ist man bei der Analyse biologischer neuronaler Netze dann in derselben Situation wie bei der Analyse künstlicher neuronaler Netze, d.h. man kann kanonisch das vor-vereinfachte Modell als dynamisches System interpretieren.

Zeitlichkeit

Eine der Grundgegebenheiten mentalen Geschehens ist seine Zeitlichkeit. In der klassischen, logik-orientierten KI, deren "Dynamik" auf intrinsisch zeitlosen Inferenzen aufbaut, kann dieser Grundtatsache nicht Rechnung getragen werden (ausführliche Diskussion und Bedeutung für das Frame-Problem in Jaeger *KI Math* 1994a). Die fundamentale Bedeutung der Zeitlichkeit für ein Verständnis kognitiver Phänomene wird von van Gelder & Port (*Kog* 1994b) zum wichtigsten Motiv für dynamische Systeme in der Kognitionswissenschaft.

Bei näherer Betrachtung löst sich die Grundtatsache Zeitlichkeit in eine kaum überschaubare Vielfalt von Phänomenen auf.

Lashley (*Bio* 1951) hat in einem aktuell anmutenden Aufsatz auf ein Rätsel hingewiesen, das bis heute noch nicht befriedigend gelöst ist: wie kommt es zur Hin- und Rücktransformation zwischen zeitlich aktualisierter Information (z.B. Sprache, Musik, Bewegungen) in statisch-räumliche (nämlich Erinnerungsspuren, die im neuronalen Substrat fixiert sind)? Lashley deutet vorsichtig die Wichtigkeit von neuronalen Rhythmen und Oszillationen an und vermutet, daß Information im Gehirn in einem Netzwerk von "reverberatory circuits" gespeichert sei. Dieser Gedanke ist heute in verschieden Versionen Allgemeingut (*Cell Assemblies* bei Hebb 1949 [zitiert nach Braitenberg 1978], kolumnäre Module bei Eccles *Bio* & Popper *Phil* 1977, B-Systeme bei Braitenberg *Bio* 1978), aber in seiner Bedeutung für das von Lashley angesprochene Rätsel immer noch nicht klar.

Viele gegenwärtige Arbeiten beschäftigen sich mit Facetten dieses Rätsels. Port et al. (*Kog*, *NN* 1994) behandeln das Problem der Erkennung zeitlicher (auditiver) Muster und finden durch experimentelle und formale Untersuchungen, daß Menschen zeitliche Muster nicht anhand einer absoluten internen Referenzzeit analysieren, sondern durch mehrere spezialisierte Verarbeitungstechniken, die u.a. periodische Regelmäßigkeiten im Signal selbst für Referenzzwecke ausnutzen.

Konkrete neuronale Mechanismen für zeitliche Charakteristika von Assoziationssequenzen anzugeben ist nicht trivial. Es werden verschiedene Mechanismen untersucht. Riedel et al. (*Phys* 1988) entwickeln eine analytische Theorie von Hopfield-Netzwerken mit zeitverzögernden Synapsen. Buhmann & Schulten (*NN*, *Phys*, *Inf* 1987) erklären Zeitfolgen von semistabilen Zuständen in asymmetrischen Hopfield-Netzwerken dagegen durch die Gegenwart von Noise. Im Kontext einer komplexen, biologisch motivierten Netzwerkarchitektur zur Musterklassifikation bauen Carpenter & Grossberg (*NN*, *Bio* 1990) zeitliche Verzögerungs- und Refraktäreffekte durch eine weitgehende Modellierung der chemischen Eigenschaften biologischer Synapsen ein. Reiss & Taylor (*NN*, 1991) nutzen lokale zeitliche Eigenschaften sog. *leaky-integrator*-Neurone in einem an Befunden aus dem Hippocampus orientierten Netz zur Speicherung von Zeitsequenzen. Ihre Arbeit enthält auch eine knappe Übersicht über andere Ansätze. Hayashi (*NN*, *Rob* 1994) benutzt ein relativ einfaches, aber chaotisches oszillatorisches Netzwerk, um Zeitsequenzen kontinuierlich ineinander überführter Muster gleichsam als "im Chaos driftenden Grenzzyklus" [intuitive Umschreibung von mir] zu lernen und zu replizieren. Yamauchi & Beer (*Rob* 1994) entwickeln mit genetischen Algorithmen rekurrente, kontinuierliche Netzwerke, welche einfache Mustersequenzen erkennen oder sich sogar ohne weiteren Gewichts-verstellenden Lernmechanismus durch Reinforcement an verschiedene solche Sequenzen adaptieren können. Leider analysieren die Autoren die erhaltenen Netzwerke nicht weiter. Die Art, wie die genetischen Algorithmen verwendet wurden, gibt jedoch Hinweise darauf, daß das neuronale Netz in Teilnetze differenziert wird, deren Aktivierungen in zeithierarchischen Beziehungen zueinander stehen.

In anderen dynamischen Systemen als neuronalen Netzen sind noch weitere Mechanismen für Aktivierungsfolgen von chaotischen Attraktoren untersucht worden. Diese Mechanismen könnten allerdings genauso auch in neuronalen Netzen auftreten. Als Beispiele sei hier auf stochastische Resonanz in chaotischen Systemen (Nicolis et al. 1993) verwiesen.

Diese Arbeiten zeigen sehr verschiedenartige Mechanismen für Zeitfolgen von Aktivierungsmustern. Sie illustrieren, daß erstens die parametrische Systemzeit t, die ein formal spezifiziertes dynamisches System besitzt, nicht notwendig schon "die" phänomenale Zeit ist, in der sich die interessanten emergenten Eigenschaften des Netzwerkes am natürlichsten interpretieren lassen, und daß zweitens zeitliche Mechanismen ein eigenständiger Forschungsgegenstand sind. Generell ist der Entwurf von (rekurrenten) Netzwerken mit gegebenen zeitlichen Verarbeitungscharakteristiken ein noch weitgehend unerschlossenes Gebiet.

Bei systemtheoretischen Analysen zeitlicher Phänomene kommt es gewöhnlich auf die Berücksichtigung verschiedener Zeitmaßstäbe für verschiedene Teilprozesse bzw. Prozeßaspekte an (z.B. Zeitkonstanten für Relaxationen gestörter Attraktoren, Parameterdrift u.a.). Einführende Bemerkungen, die ein Gefühl für diese Dinge vermitteln, finden sich bei Kelso et al. (*Phys*, *Bio*, *Psych* 1993) und Jaeger (*KI*, *Math* 1994c), ein Beispiel in Schöner et al. (*Phys*, *Bio*, *Psych* 1986).

Als letzte Arbeit sei auf Elman (*Ling*, *Kog* 1990) verwiesen. Er konstruiert ein rekurrentes lokalistisches Netzwerk, das einen endlosen Inputstrom von Wörtern erhält, aus darin implizit enthaltenen zeitlichen (genauer: Ordnungs-) Regularitäten unüberwacht grammatische Strukturen und hierarchisch geordnete Wortklassen lernt, und dann in "Echtzeit" ähnliche Inputströme analysiert. Diese Arbeit ist zwar weder formal systemtheoretisch orientiert, noch strebt sie biologische Plausibilität an. Sie ist aber ein (vielzitiertes) Beispiel für den erfolgreichen Einsatz rekurrenter Netzwerke nach dem Motto "Mut zur Dynamik" in einem traditionell strukturell-grammatikalisch denkenden Fachgebiet.

Die von van Gelder & Port (1994b) eingeforderte Untersuchung von Zeitlichkeit ist notwendig für ein Verständnis der physikalischen Situiertheit intelligenter Agenten. Die in der Robotik- und Autonome-Agenten-Literatur oft zu findende Forderung nach "Echtzeitverhalten" verkennt dabei wohl noch die Tiefe und phänomenalen Reichhaltigkeit des Problems. Dynamische Systeme sind zwar mit einer "eingebauten Echtzeit" t ausgestattet. Die Frage nach der Zeitlichkeit des Mentalen wird dadurch aber beileibe nicht schon beantwortet, sondern t stellt nur ein Mittel zur Verfügung, um überhaupt zweitliche Phänomene beschreiben zu können.

Zeitliche Periodizitäten

Eine klassischer Gegenstand systemtheoretischer Analysen ist periodisches Verhalten in allen möglichen Varianten (schöne Einführung: Abraham & Shaw *Math* 1983, Band 1, sehr komprimiert auch in Kelso et al. *Math Psych* 1993). Periodische Phänomene finden sich in neuronalen Netzen auf vielen Ebenen und werden mit den Methoden dynamischer Systeme untersucht.

Gekoppelte Oszillationen sind für das Verständnis der neuronalen Mechanismen kognitiver Informationsverarbeitung möglicherweise fundamental wichtig. Dem Konnektionismus wurde einst vorgeworfen (Fodor & Pylyshin *Phil Kog* 1988), mit dem Phänomen der Kompositionalität kognitiver Verarbeitungseinheiten (*variable binding*, *feature linking*) nicht fertigzuwerden. Dieser Angriff induzierte in kurzer Zeit die Entdeckung einer ganzen Reihe von konnektionistischen Mechanismen zur Kompositionalität (vgl. van Gelder *Kog* 1990, Chalmers *Phil Kog* 1992, van Gelder & Port *Kog* 1994a). Ein heute bevorzugt untersuchter Mechanismus besteht nun gerade in global gekoppelten neuronalen Oszillationen, d.h. Synchronisierungen zwischen mehr oder weniger voneinander entfernten Neuronen.

Solche Synchronisationen wurden experimentell in biologischen Netzen nachgewiesen, wo sie mit der Verknüpfung einzelner sensorischer Features in Verbindung gebracht werden konnten (Engel et al. *Bio* 1990, Eckhorn & Stoecker *Bio* 1994). Stoecker & Eckhorn (*Bio* 1994) modellieren die empirischen Befunde in einem relativ einfachen neuronalen Netz in der biokybernetischen Tradition. Grossberg & Somers (*NN*, *Bio*, *Rob* 1991) konstruieren für einen erweiterten Aufgabenbereich ein komplexes "biologisch plausibles" Modell eines muster-erkennenden Netzwerkes, wo ebenfalls synchronisierte Oszillationen für die Kopplung einzelner Reizkomponenten sorgen. In dieser Arbeit werden einige Varianten des Kopplungsmechanismus in der typischen Mischung aus formaler Analyse und Simulationsstudien untersucht. Ein Vorzug des Synchronisierungsprinzips besteht darin, daß die Kopplung von sensorischen Features kein isoliertes Phänomen ist, sondern flexibel in einen "ganzheitlichen" Wahrnehmungsapparat eingebunden ist und z.B. durch Aufmerksamkeits-prozesse beeinflußt wird. Die Arbeit steht vor dem Hintergrund der *Adaptive Resonance Theory* (ART), die inzwischen eine ganze Familie von Netzwerkarchitekturen für Klassifika-tionsaufgaben hervorgebracht hat, welche auch von anderen Forschern verwendet werden.

Als letzter Vertreter des Prinzips Kompositionalität durch Synchronisation seien Mani & Shastri (*Kog*, *KI* 1993) genannt. Im Unterschied zu den meisten anderen hier referierten Arbeiten verwenden sie lokalistische Netzwerke. Dies ist typisch für konnektionistische Techniken in den Kognitionswissenschaften und der Linguistik. Die Autoren präsentieren eine komplexe Netzarchitektur für Aufgaben des regelbasierten hierarchischen Klassifizierens. Variablenbindungen werden durch Synchronisation von *spike trains* geleistet. Im Gegensatz zu wohl allen bekannten klassischen KI-Systemen für denselben Zweck besitzt ihr Netzwerk im Prinzip (bei paralleler Implementierung) eine lineare Zeitkomplexität in der Länge der Eingabe, und ist beliebig erweiterbar ohne Beeinflussung der Verarbeitungsgeschwindigkeit. Entsprechend ihrem disziplinären Hintergrund bieten Mani & Shastri keine systemtheoretische formale Analyse ihres Netzwerks. Ihre Arbeit wurde dennoch hier erwähnt, weil sie den Nutzen einer dynamischen, d.h. prozeßorientierten Sicht auf scheinbar strukturelle, klassische KI-Phänomene (hier: Kompositionalität) demonstriert.

Einige Phänomene der Wahrnehmung von Rhythmen modellieren Large & Kohen (*Psych* 1995), indem sie formal neuronale Module als phasen- und frequenzschließende Oszillatoren spezifizieren. Diese vorbildlich klar geschriebene und sehr anregende Arbeit enthält viele Literaturhinweise zum Themenkomplex gekoppelter Oszillatoren, neuronaler Modellierung von Rhythmen, und Wahrnehmung temporaler Muster, und kann auch als Einführung in dies Gebiet gelesen werden.

Informationsspeicherung bzw. -repräsentation

Neuronale Netzwerke als verteilte Informationsspeicher haben einige eingebaute Vorzüge wie *graceful degradation*, Generalisierung und Vervollständigung von gepeicherten Mustern, Lernfähigkeit u.a. (Einführung in verschiedene Repräsentationstechniken: Dorffner *KI*, *NN* 1994, zu den inzwischen klassischen "Assoziativspeichern" vgl. z.B. Palm *NN*, *Math*, *Bio* 1988). Im Rahmen dieser Übersicht sind unter der Vielzahl von untersuchten Speicher- bzw. Repräsentationsprinzipien vor allem solche interessant, wo die Information in *Prozessen* repräsentiert ist, die beim Abruf (re-)aktiviert werden. Hierbei versagt die traditionelle Vorstellung der Informatik, wonach Speicherung und Abruf auf der einen Seite und die Verarbeitung von Information auf der anderen auseinanderzuhalten sind. Vielmehr sind die Abrufprozesse schon gleich die Verarbeitungsprozesse; der dynamische Speicher arbeitetselbst (Auswirkungen auf das *symbol grounding* Problem diskutiert Chalmers *Phil*, *Kog* 1990).

Seit einigen Jahren werden im Konnektionismus die Funktion *chaotischer* neuronaler Prozesse bei der Informationsspeicherung untersucht. Yao & Freeman (*Bio*, *NN* 1990) rekonstruieren anhand physiologischer Befunde das Riechhirn (ursprünglich von Kaninchen) formal als dynamisches System und untersuchen die Dynamik in Simulationsstudien. Das Hauptresultat besagt, daß das Riechhirn im Grundzustand sich in einem hoch chaotischen Attraktor befindet und bei Reizvorgabe in weniger chaotische oder Grenzzykel-nahe "Flügel" dieses Attraktors stabilisiert. Diese Flügel sind also dynamische Repräsentationen bestimmter sensorischer Reize. In Yao et al. (*NN*, *Rob* 1991) demonstriert dieses Modell in einer industriellen Anwendung seine Robustheit. Ähnlich wird die Bedeutung eines chaotischen Grundzustandes bei der Informationsspeicherung und dem Retrieval auch von Babloyantz & Lourenço (*Bio*, *NN* 1994) herausgestellt. Sie weisen darauf hin, daß ein chaotischer Attraktor im Prinzip unendlich viele semistabile Teilmuster enthält, also ein im Prinzip unbegrenzte Speicherkapazität besitzt, und sie interpretieren den chaotischen Grundzustand im Sinne eines Aufmerksamkeitszustandes. Dies letztere wird wiederum ähnlich in der bereits zitierten Arbeit von Hayashi (*NN*, *Rob* 1994) gesehen, der einen chaotischen Grundzustand als "Suchzustand" interpretiert. Happel & Murre (*Psych*, *NN* 1995) beschreiben modulare rekurrente Netzwerke, die ohne Reinforcment einen zweidimensionalen Inputraum klassifizieren lernen. Es entwickelt sich eine reiche chaotische Dynamik, die zu fraktalen Klassengrenzen führen. Die Autoren diskutieren die Funktion einer chaotischen Dynamik unter verschiedenen Aspekten, u.a. als "novelty filter", als Suchzustand, und als Mechanismus zur Bildung komplexer Klassengrenzen. Schließlich möchte ich hier noch Tani (*Inf* 1995) aufführen, der ein rekurrentes neuronales Netz Symbolsequenzen lernen läßt, die von einem nichtdeterministischen endlichen Automaten geliefert werden. Die Lerngeschichte ist eine Feigenbaum-Kaskade von Bifurkationen, und das erfolgreich trainierte Netz zeigt einen globalen chaotischen Attraktor, dessen Struktur sehr schön mit der Struktur des Stimulus-generierenden endlichen Automaten in Zusammenhang gebracht werden kann. Das Netzwerk generiert "Symbole", indem die Trajektorie bestimmte Zellen des Phasenraumes durchläuft.

Man ersieht aus diesen Arbeiten, daß die Hebb'sche Grundidee von ko-aktivierten Zellverbänden als Grundeinheit der Informationsrepräsentierung noch sehr lebendigen Forschungsstoff abgibt!

Kommentar

Die Themen Zeitlichkeit, Periodizität und Speicherung/Repräsentation sind eng miteinander verwandt. Die meisten der zitierten Arbeiten hätten in allen drei Abschnitten untergebracht werden können. Als übergeordnetes Thema ist vielleicht "dynamische Repräsentationen" zu sehen. Fragen der Zeitlichkeit ergeben sich zwangsläufig mit dem Problem der Abfolge einzelner repräsentierender Netzzustände, und Periodizitäten kommen einerseits durch die Hebb'sche Grundidee von positiv rückgekoppelten neuronalen Verbänden und andererseits durch Synchronisierungsmechanismen zur Kompositionalität ins Spiel — beides formal *Resonanz*phänomene (vgl. Grossberg & Somers *NN*, *Bio* 1991, Smolensky *NN*, *Kog* 1986). Besonders schön wird die Verquickung der drei Aspekte in der einfachen Netzarchitektur von Matsuga & Yuille (*NN*, 1994) deutlich. Sie beschreiben die Repräsentierung von raumzeitlichen Mustern, wobei sie Neurone mit autonomem periodischem Spikeverhalten ausnutzen. Auch in den vorgestellten komplexen, biologisch orientierten Netzwerken (Yao &

Freeman 1990, Yao et al. 1991, Carpenter & Grossberg 1990, Grossberg & Somers 1991) sind alle drei Aspekte eng miteinander verbunden.

Das Thema "Lernen" ist bekanntlich in der Neuronale-Netz-Forschung zentral. Es fehlt hier deshalb, weil Lernvorgänge im Konnektionismus kaum mit systemtheoretischen Techniken analysiert werden. So sind mir nur sehr wenige Arbeiten bekannt (Tani *NN* 1995, Happel & Murre *NN* 1995), in der Lernfortschritte eines neuronalen Netzes als eine Sequenz von Bifurkationen analysiert werden, was systemtheoretisch naheliegend wäre (vgl. Abschnitt 4.2.2). Es ist zwar zu Beginn dieses Abschnitts die kanonische Deutbarkeit von neuronalen Netzen als dynamische Systeme herausgestellt worden. Das Thema "Lernen" scheint aber gerade der eigenständige, nicht direkt auf bekannte systemtheoretische Konstrukte abbildbare Anteil im Konnektionismus zu sein.

Hier ist außerdem anzumerken, daß die prominentesten "Lern-Erfolge" im Konnektionismus vielfach auf Varianten des Backpropagation-Algorithmus' (Einführung: Rojas *NN*, *Math* 1994) und der *self-organizing feature maps* (Obermayer et al. *NN* 1990a) beruhen. Beides betrifft Netzwerke, die nicht (vollständig) rekurrent sind, d.h. die keine autonome Aktivität aufrechterhalten können. Bei solchen Netzwerken sind systemtheoretische Methoden aus der synergetischen und der differentialgeometrischen Tradition nicht kanonisch anwendbar, wohl aber Methoden der ingenieurswissenschaftlichen Systemtheorie. Denn in der ingenieurswissenschaftlichen Tradition ist ein System durch sein Ein-Ausgabeverhalten bestimmt, im einfachsten Fall eine "gedächtnislose" Funktion. Dies trifft auf feed-forward-Netzwerke zu. Tatsächlich gibt es viele Arbeiten, die sich mit feed-forward-Netzwerken als Funktionsapproximatoren beschäftigen. So geben Leshno et al. (*NN*, *Math* 1993) eine hinreichende und notwendige Eigenschaft für feed-forward-Netze an, die beliebige kontinuierliche Funktionen approximieren können. Caelli et al. (*Rob*, *NN* 1993) geben ein Verfahren an, wie man feed-forward-Netze mit vorgegebenen Filtereigenschaften konstruiert — ein Beispiel von Ingenieurskunst mit kontrolltheoretischen Methoden.

Abschließend soll noch ein Punkt herausgestellt werden, der vielleicht einem tiefsitzenden Vorurteil von seiten der klassischen KI widerspricht. Naturwissenschaftlich und differentialgeometrische Theorien dynamischer Systeme beschäftigen sich mit *qualitativen* Eigenschaften von Systemen, wie Attraktoren, Bifurkationen etc. Dem entsprechen in neuronalen Netzwerke "höhere" Phänomene: Aktivierungsmuster als ganze, Folgen davon, chaotische globale Zustände, etc. Es geht im Konnektionismus zumeist gar nicht um die elementare "hardware"-Ebene im neurologischen Detail (dazu sind künstliche neuronale Netze auch viel zu sehr vereinfachende Abstraktionen), sondern um globalere Funktionsprinzipien. Genauso geht es in der logikorientierten KI ja auch nicht etwa um die Untersuchung der elementaren logischen Junktoren selbst, sondern um das, was man durch die Verwendung sehr vieler davon in einem größeren Maßstab ausrichten kann. Kurz, die Abstraktionsebene im Konnektionismus ist höher als aus der Sicht der KI wohl wahrgenommen wird.

4.2.2 Sensomotorik

Sich bewegende physikalische Körper (Planeten, Pendel) sind die frühesten mathematisch analysierten dynamischen Systeme (überraschende historische Anmerkungen in Abraham & Shaw *Math* 1983f Band 1 und 4). Menschen und andere situierte Agenten haben bewegliche physikalischen Körper. So ist es kein Wunder, daß "harte" systemtheoretische Techniken bereits seit längerem in der Biophysik und der psychologischen Motorikforschung ihren Einzug gehalten haben. Es sind dort Einsichten gesammelt worden, die für die KI und auch für die Theorie autonomer Agenten von großem Interesse sind. Der systemtheoretische Hintergrund ist zumeist die Synergetik.

Eine anderer, ebenfalls hier darzustellender Zusammenhang zwischen Motorik und Systemtheorie besteht durch die klassische Robotik, welche sich u.a. mit der ingenieursmäßige

Regelung von Effektorbewegungen und Navigationsaufgaben beschäftigt, was beides die Verarbeitung von Sensordaten einschließt. Das Vorgehen ist modellbasiert, d.h. der erste Schritt bei der Entwicklung eines Bewegungskontrollsystems besteht in der formalen Modellierung des zu kontrollierenden Effektors, evtl. sogar der dazugehörigen Sensorcharakteristika und der Umwelt. Der systemtheoretische Hintergrund ist hier die ingenieurswissenschaftliche Systemtheorie, Kybernetik und Kontrolltheorie.

In neueren Arbeiten werden beide Traditionen auch miteinander verbunden, z.B. in selbstorganisierenden Neurokontrollern für Roboter. Für die KI ist diese Verbindung von ingenieursmäßigem Entwurf, Selbstorganisationsfähigkeiten und synergetischer Analyse inspirierend.

Bewegungsmuster

Biologische Bewegungsapparate haben ungeheuer viele Freiheitsgrade, d.h. physikalisch voneinander mehr oder weniger unabhängig, die Bewegung bestimmende Größen. Dazu gehören Winkelstellungen und -geschwindigkeiten an Gelenken, Raumkoordinaten und Geschwindigkeiten von Gelenken, Kräfte an einzelnen Sehnen, bei feinerer Modellierung auch Kräfte einzelner Muskelfasern oder sogar Aktivierungen einzelner Motoneurone. Mit klassischen kontrolltheoretischen Methoden läßt sich ein solches System weder theoretisch modellieren noch praktisch regeln, und so war es lange rätselhaft, wie dies biologische Systeme schaffen (Smith & Thelen *Psych* 1993b). Hier bietet die Synergetik mit dem Prinzip der Versklavung eine Erklärungsmöglichkeit an. Im Kern besagt sie, daß die vorhandenen Kopplungen die einzelnen physikalischen Einflußgrößen zu einem selbstorganisierten System zusammenbinden, in dem wenige Ordnungsparameter die große Anzahl von physikalischen Freiheitsgraden versklaven. Ein solches System kann dann auf der Ebene der Ordungsparameter beschrieben bzw. durch deren Variation reguliert werden.

Diese Idee wurde zur Grundlage einer inzwischen wohl klassisch zu nennenden Studienserie zur Analyse periodischer Handbewegungen (Schöner et al. *Phys Psych* 1986, einführende Beschreibung in Kriz *Psych* 1992, 150ff). Versuchspersonen wurden aufgefordert, mit den Händen gegenläufige periodische Bewegungen vorgegebener Frequenz auszuführen. Bei Steigerung der Frequenz kommt es zu einem unwillkürlichen Umschlag in eine gleichläufige Bewegung. Es zeigte sich, daß zur Bestimmung dieses Systems ein einziger Ordnungsparameter ausreicht, nämlich die Phasenverschiebung zwischen den beiden Bewegungen. Er ergibt sich aus einer mathematischen Modellierung des Systems durch zwei nichtlinear gekoppelte nichtlineare Oszillatoren. Bei mathematischer Berücksichtigung von stochastischen Fluktuationen können die empirisch gefundenen stochastischen Charakteristika des Bewegungsumschlages im Detail erklärt werden. Der Bewegungsumschlag selbst ist eine Bifurkation zwischen verschiedenen Attraktoren.

In Smith & Thelen (*Psych* 1993a) finden sich mehrere Aufsätze, in denen motorisches Verhalten des Menschen aus naturwissenschaftlich-systemtheoretischer Sicht analysiert wird. Zwei davon seien hier angeführt. Clark et al. (*Psych* 1993) beschreiben Laufbewegungen als Grenzzyklus. Durch verschiedene experimentell induzierte Störungen weisen sie die strukturelle Stabilität des Bewegungsmusters nach und beschreiben Ober- und Unterschenkel als System gekoppelter Oszillatoren. Bertenthal & Pinto (*Psych* 1993) arbeiten die These aus, wonach Menschen die Bewegungen anderer Menschen durch organisierende Prinzipien wie periodische Attraktoren, Phasenschluß u.a. wahrnehmen. Durch Experimente untermauern sie die Vermutung, daß die Ordnungsparameter, welche bei der visuellen Wahrnehmung von Bewegung entscheidend sind, dieselben sind wie die für die Generierung derselben Bewegungen wirksamen. Dies hilft, Bewegungslernen durch Imitation zu erklären. Robertson et al. (*Psych* 1993) beschäftigen sich mit den spontanen Bewegungen des Neugeborenen. Aus empirischen Bewegungstraces rekonstruieren sie einen chaotischen Attraktor, dessen fraktale

Dimension sie abschätzen. Diese ist niedrig (etwa 3.3), was bedeutet, daß das System durch einen ebenfalls niedrigdimensionalen Zustandsraum erfaßt werden kann, also im in gewissem Sinne "einfach" ist. Die Autoren verwenden etwas knifflige und nicht vollkommen verstandene numerische Techniken (knapp im systemtheoretischen Kontext dargestellt bei Nicolis *Phys* 1989, relevante kritische Bemerkungen bei Preissel et al. *Bio* 1990, Millonas *Math*, *Phys* 1994). Die Schwierigkeiten bestehen im Kern darin, bei Datenmaterial von geringem Umfang Noise von deterministischem Chaos zu trennen. Eine formal völlig analoge Analyse wurde von Smithers (unveröffentlichter Vortrag) für die Bewegungen eines Roboters in einer einfachen Arena durchgeführt.

Wolff (1993) diskutiert verschiedene Wach- und Schlafzustände beim Kleinkind. Er interpretiert sie als Attraktoren in der Interaktionsdynamik vieler, nicht näher spezifizierter physiologischer und motorischer Subsysteme. Er belegt diese Auffassung u.a. durch detaillierte Beschreibungen von Übergängen zwischen Zuständen, die für Phasenübergänge typische Merkmale aufweisen. Die Ausbildung einer Kopplung von emotionalen Zuständen und deren motorischem Ausdruck, die interkulturell wenig variiert, führt er nicht auf eine spezielle genetische Steuerung zurück, sondern auf allgemeine Prinzipien der Selbstorganisation.

Eine Beschreibung von Bewegungen in einem ganz anderen, aber ebenfalls systemtheoretisch motivierten Sinn findet man bei Townsend & Busemeyer (*Psych* 1989). Sie modellieren Appetenz-Vermeidungs-Konflikte durch Trajektorien in Kräftefeldern. In der Robotik verfolgen Schöner & Dose (*AA* 1992) einen analogen Ansatz. Sie beschreiben den Wegfindeprozeß eines (allerdings nur simulierten) mobilen Roboters als das Verfolgen von Potentiallinien in einem durch Differentialgleichungen spezifizierten Kräftefeld. Dieser Ansatz ist weiterentwickelt in Engels & Schöner (*AA* 1995), wo eine modulare Architektur eines wegfindenden Roboters entworfen wird. Die einzelnen Module sind formal gleichartige, dynamische skalare Felder. Durch geeignete Paramterwahl kann die Dynamik jedoch einen unterschiedlichen Typus annehmen (z.B. Entscheidungsfällung, Kurzzeitgedächtnis, Langzeitgedächtnis).

Kontrolle von Bewegungen

In der Robotik ist die Regelung von Effektorbewegungen und räumlicher Fortbewegung eine Grundaufgabe, die neben fuzzy-logic-Methoden vor allem mit kontrolltheoretischen und konnektionistischen Techniken angegangen wird. Fast immer ist die Bewegungskontrolle mit der dualen Aufgabe der Sensordatenanalyse gekoppelt, die in diesem Abschnitt implizit mit vorgestellt wird. Es kann nur eine illustrative Probe aus der umfangreichen Literatur vorgestellt werden.

Ein beeindruckendes Beispiel eines klassischen ingenieursmäßigen Ansatzes bietet Triggs (*Rob* 1994). Er beschreibt die Naviagationssteuerung eines mobilen Roboters auf der Grundlage von Sonarsensoren. Diese haben eine geringe Winkelauflösung und sind anfällig für Fehlinterpretationen durch Mehrfachreflexionen. Diese Schwierigkeit wird weitgehend bewältigt durch eine detaillierte mathematische Modellierung der geometrischen, physikalischen und stochastischen Eigenschaften der verwendeten Sensoren und der häufigsten Umweltelemente (!). Dies Modell dient dann als Grundlage für Kalman-Filter zur stochastisch optimalen Auswertung der Sensordaten. Der Aufwand wird als beträchtlich geschildert. Die Motivation für einen solchen rein modellbasierten, heute eher unüblich klassischen Ansatz besteht für den Autor darin, daß er am Ende ein klares mathematisches Modell des Vorganges hat.

Steels (*KI*, *AA* 1994b) rekonstruiert einige Behaviors in seinem mobilen Roboter als Punktattraktor und verwendet diese Formalisierung als Korrektheitskriterium für die Programmierung der an dem Verhalten beteiligten Prozesse. Kawato (*Rob* 1994) beschreibt ein Kontrollsystem für Effektorbewegungen, das hierarchisch von der Trajektorienplanung bis zu den Motorkommandos absteigt und aufsteigende Informationsflüsse integriert (ältere, aber lesenswerte Einführung in Kontrolltheorie hierarchischer Systeme: Powers *Rob* 1988). Da (zumindest biologische) Sensoren zu träge und ungenau, und Signalleitungsgeschwindigkeiten zu langsam sind, um eine klassische feedback-Kontrolle der Bewegung zu unterstützen, muß die Bewegungskontrolle "vorausschauend" sein, d.h. sie muß ein (inverses kinematisches und dynamisches) Modell des Bewegungsapparates beinhalten. Die notwendigen top-down Berechnungen werden in Kawatos Architektur durch bottom-up feedback beschleunigt. Dieser Informationsfluß benötigt Vorwärts-Modelle der Kinematik und Dynamik. Dieser komplexe, aber klassisch kontrolltheoretische Entwurf wird durch neuronale Netze verwirklicht. Schweitzer & Wen (*Rob* 1994) stellen fest, daß neuronale Netze in der Robotik zwar in Mode, aber kein Allheilmittel für schwierige Kontrollaufgaben sind, und präsentieren einige Fälle, wo nach ihrer Meinung der Einsatz dieser Technik sinnvoll ist. Unter anderem beschreiben sie einen Ping-Pong spielenden Roboter, bei dem der systematische Fehler eines klassischen modellbasierten Kalmanfilters durch ein neuronales Netz ausgeglichen wird.

Die kontrolltheoretische Sicht auf die Bewegungssteuerung ermöglicht eine funktionales Interpretation biologischer neuronaler Strukturen. Ein schönes Beispiel ist die Studie von Ewert und von Seelen (*Bio* 1973) über die Verarbeitung zweier Klassen reflexauslösender, visueller Stimuli bei Kröten. Hier werden neurobiologische und experimentelle Befunde in ein formales Modell im Sinne der Kontrolltheorie abgebildet.

Viele Arbeiten beschäftigen sich mit der Funktion des Kleinhirns, da es anatomisch besonders einfach und gut untersucht ist, und außerdem seine Funktion klarer ist als bei anderen Gehirnteilen. Es wird als der Ort angesehen, in dem Modelle des Bewegungsapparates und Kalman-Filter zur Trajektorienverfolgung liegen (Kawato *Rob* 1994, Paulin *Bio* 1993). Gaudiano & Grossberg (*Bio, Rob, NN* 1991, p. 179f) geben detailliert an, wo die vielen Komponenten ihres AVITE-Modells (s.u.) der Bewegungskontrolle im Gehirn realisiert seien. Das neuronale Netz zur Bewegungskontrolle sechsbeiniger Insekten, welches Cruse et al. (*Bio* 1994) beschreiben, ist das Ergebnis einer Kombination kontrolltheoretischer Modellbildungen und neurobiologischer Befunde von Stabheuschrecken.

Zusammenfassend läßt sich feststellen, daß hier eine reiche, auf die kybernetische Tradition zurückgehende Forschungslandschaft existiert, in der kontrolltheoretische, neurobiologische und konnektionistische Methoden die biologische Bewegungskontrolle aufklären helfen und zu künstlichen Kontrollsystemen für die Robotik führen. Diese Forschung steht dennoch erst am Anfang. Das wird im nächsten Abschnitt klarwerden, wo auf Fragen der Adaptivität, der Entwicklung und des Lernens von motorischem Verhalten eingegangen wird.

Lernen und Entwicklung von motorischem Verhalten

Menschen zeigen in ihrer Lerngeschichte typischerweise auffällige Diskontinuitäten. Plötzliche qualitative Sprünge wechseln sich mit längeren Perioden scheinbarer Stagnation oder langsamer Reifung ab. Dieser Befund gilt für alle in der Entwicklungspsychologie betrachteten Kompetenzen, unter anderem für den Spracherwerb und das Erlernen von Bewegungsmustern. Zur Erklärung dieses Phänomens wurden in der Psychologie komplexe neuronale oder mentale, durch genetische Programme zur Reifung gelangende Mechanismen für jede neu auftauchende Einzelfähigkeit oder Fertigkeitsstufe postuliert (Smith & Thelen *Psych* 1993c, Goldfield *Psych* 1993). Aus systemtheoretischer Sicht bieten sich Erklärungsmöglichkeiten an, die auf sehr allgemeinen Prinzipien nichtlinearer Systeme beruhen und keine speziellen Mechanismen stipulieren müssen. Es gibt zur Zeit wohl zwei Ansätze, nämlich die Deutung von qualitativen Entwicklungssprüngen als Bifurkationen und die Erklärung von Diskontinuitäten durch nichtlineare Interaktionen zwischen Teilfähigkeiten.

Ein schlagendes Beispiel für die erste Art der Erklärung gibt Thelen (*Psych* 1984, referiert nach Goldfield *Psych* 1993) an. Ihr gelang es, das bei Neugeborenen nach 8 Wochen zu beobachtende Verschwinden eines Trittreflexes auf die einfache Tatsache zurückzuführen, daß das Bein Fett ansetzt, im Verhältnis zur Muskelkraft schwerer wird, und die ursprüngliche Trittbewegung aus rein "oszillationsmechanischen" Gründen kein stabiles Bewegungsmuster mehr ist. Goldfield (*Psych* 1993) analysiert gründlich das Krabbelverhalten bei Kleinkindern. Es gibt viele Formen des Krabbelns, die teils nacheinander, teils gleichzeitig auftreten, bei verschiedenen Kindern verschieden. Das Gesamtbild, das Goldfield zeichnet, besagt, daß die einzelnen Krabbelformen, angestoßen durch partikuläre Umwelterfordernisse, aus den gerade zur Verfügung stehenden neuralen und skeletomuskulären Elementen des Bewegungsapparates zusammengestellt werden. Wenn sich einzelne Elemente verändern, kann dies zu Bifurkationen, d.h. Reorganisationen des Bewegungsmusters, führen. Diese Sicht betont die Bedeutung von Umwelt*affordances*, Fluktuationen und Attraktoren, die "opportunistisch" eingenommen werden, ohne daß spezifische Mechanismen involviert sind.

Ein anderer Erklärungsansatz für offenbare Diskontinuitäten in der Entwicklung besteht darin, verschiedene Teilfertigkeiten als nichtlinear miteinander gekoppelte und um Ressourcen (wie etwa Aufmerksamkeit) konkurrierende "Populationen" zu interpretieren. Dieser Weg wird von van Geert (*Psych* 1993) in detaillierten Rekonstruktionen empirischer Zeitserien besonders aus dem Gebiet des Spracherwerbs beschritten. Die klassischen systemtheoretischen Gleichungen populationsdynamischer Prozesse liefern, je nach Parametern und verschiedenen anderen Details, eine Vielzahl von Wachstumskurven, die Phänomen wie den "plötzlichen" Durchbruch einer Fertigkeit, Oszillationen, kurzzeitige Abschwächung einer Fertigkeit vor einem steilen Anstieg u.a.m. erfassen. Van Geert geht ausführlich auf methodologische Fragen ein und kommt zu dem Schluß, daß systemtheoretische Techniken der empirischen Entwicklungsforschung wesentlich feiner auflösende Modelle und damit präzisere Prädiktionen zur Verfügung stellen, als das mit den herkömmlichen (Gruppen-)statistischen Mitteln möglich ist.

Der Unterschied zwischen der Erklärung durch Bifurkationen und dem Ausnutzen nichtlinearer Interaktionen besteht darin, daß Bifurkationen das Entstehen qualitativ neuer Fertigkeiten modellieren, während Nichtlinearitäten quantitative Entwicklungsschübe einer schon bestehenden Fertigkeit erklären.

Die hier umrissene neue systemtheoretische Sicht der Psychophysik stellt zwar einen große Vereinfachung und Vereinheitlichung von Erklärungen einzelner Bewegungen und deren Entwicklung dar. Es bleibt jedoch zu klären, welche Konstitution der sich entwickelnde Organismus anfangs aufweisen muß, damit die Selbstorganisation tatsächlich zu einer gelungenen Anpassung an die Umwelt führt. Auf die Robotik bezogen lautet diese Frage: Wie sind Roboter zu konstruieren, die sich nach ihrer Freisetzung autonom an die Umgebung anpassen und sogar zunehmend höhere Verhaltensweisen entwickeln sollen?

Wesentliche Punkte bei diesem Problem sind heute noch unklar. Das Lernproblem, das hier angesprochen ist, ist erschwert durch eine ganze Reihe von Faktoren:

- Es sollen nicht nur vom Konstrukteur mitgegebene Basisfähigkeiten optimiert, sondern ganze neue, höhere Verarbeitungsebenen erzeugt werden.
- Der Lernprozeß darf die Überlebensfähigkeit des Agenten nicht beeinträchtigen (Steels *KI, AA* 1994c)
- Es ist nicht klar, nach Maßgabe welcher Ziele der Agent seine Entwicklung optimieren soll, d.h. welches Kriterium er zur Bewertung von Verhaltensalternativen verwenden soll, oder noch anders ausgedrückt, wie das Reinforcement für die Kontrolle des unüberwachten Lernens entstehen soll. Die Frage ist reichlich dunkel (vgl. Steels *KI, AA* 1994a, Cabanac *Psych* 1992, Jaeger *KI, Math* 1994a, p. 45ff).
- Es besteht möglicherweise ein inhärenter Widerspruch in der Forderung, einen Agenten zu *konstruieren,* der sich *autonom* entwickelt. Ich vermute, daß die Ko-Evolution von Agent und Umwelt bei biologischen Agenten eine Bedingung für deren Adaptationsfähigkeit ist.

Eine solche Ko-Evolution wäre aber durch einen Konstruktionsakt für künstliche Agenten nur schwer einzuholen.

Trotz dieser Schwierigkeiten sind einige Ansätze zur autonomen Entwicklung situierter Agenten entwickelt worden. Sie lassen sich aufteilen in solche, bei denen *evolutionäre* Mechanismen für die Entwicklung qualitativer Neuerungen sorgen, und solche, in denen keine solche Mechanismen vorgesehen sind, jedenfalls nicht auf direkt ersichtliche Weise.

Ein inzwischen recht verbreitete Methode besteht darin, neuronale Netze, die zur Kontrolle eines Agenten eingesetzt werden sollen, off-line in Simulationen mit genetischen Algorithmen zu generieren oder zu optimieren. Auf diese Weise gewinnt z.B. Beer (*AA* 1995) einen neuronalen Controler für die Beinwegungen einer Laufmaschine. Diese Vorgehensweise kann nicht für die Adaptation eines einzelnen, physikalisch realisierten Agenten übertragen werden, der während des Vorganges "lebensfähig" bleiben muß. Sie soll daher in diesem Zusammenhang nicht weiter betrachtet werden.

Einen interessanten Zwischenbereich zwischen Simulation und physikalischer Realität besetzen Harvey et al. (*AA*, *Rob* 1994). Sie bauen physikalisch den optischen Apparat eines zu entwickelnden Agenten und optimieren ihn isoliert in einer optisch einfach strukturierten physikalischen Umgebung, wobei die Bewegung des Agenten durch Bewegungen der Aufhängung des optischen Apparats simuliert werden. Zur Optimierung werden genetische Algorithmen verwendet. Diese Methode kombiniert die "Echtheit" einer physikalischen Sensor-Umgebungs-Situation mit der Replizierbarkeit und teilweise auch der Geschwindigkeit von simulierten Agenten, welche mit der Verwendung evolutionärer Verfahren harmoniert.

Steels (*KI*, *AA* 1994c) schließlich liefert ein Beispiel zur "on-line Evolution" von behavior-basierten Robotern. Aus zufälligen Populationen von sensomotorischen Elementarprozessen evolvieren Prozeßkombinationen (*behavior systems*), die vorgegebene einfache Bedingungen (z.B. "maximale Geschwindigkeit") verwirklichen. Der Mechanismus ist schnell (Größenordnung 30 Generationen bis zur Etablierung einer erfolgreichen Prozeßkombination) und findet "on-line" auf dem operierenden Roboter statt. Der Prozeß ist allerdings insofern nicht wirklich evolutionär, als er ein vorgegebenes Ziel ansteuert.

Es existiert gegenwärtig wohl kein Beispiel, wo ein physikalisch realisierter Roboter durch evolutionäre Mechanismen zu einer Entwicklung oder zumindest Adaptivität, die den Namen verdient, verholfen wird. Nach meiner Meinung ist die hier auftretende Frage nach der grundsätzlichen Möglichkeit einer solchen Strategie unbeantwortet. Dagegen spricht, daß evolutionäre Mechanismen primär auf der Abfolge von Generationen verschiedener Individuen beruhen, während Adaptation sich innerhalb eines einzelnen Individuums abspielt. Nun kann man die Beschreibungsebenen drastisch verschieben und die vielen Prozesse, Konzepte oder Fähigkeiten *innerhalb* eines Agenten als *Arten* ansehen, die während der Lebensdauer desselben eine Evolution durchmachen. Der Agent wird hier als "Ökosystem" aufgefaßt, welches vielerlei Arten umfaßt (z.B. von van Geert *Ling* 1993, der diese Metapher explizit ausbaut und die Entwicklung linguistischer Kompetenzen mit denselben Gleichungen beschreibt, wie sie in der elementaren Populationsdynamik verwendet werden; von Steels *AA* 1994c; oder bei Shimohara *Inf* 1994, der grandiose Zukunftsprojekte zur Software- und Hardwareevolution künstlicher Gehirne entwirft). In dieser Perspektive macht man zumindestens keinen offenbaren kategorialen Fehler, wenn man die Evolutionsmetapher auf die Entwicklung eines einzelnen Agenten anwendet. Meiner Meinung nach führt diese Perspektive gleichwohl ins Abseits, da Agenten so etwas wie Identität, Integrität oder Individualität besitzen sollten, Ökosysteme aber grundsätzlich "fließend" sind und sich im Laufe evolutionärer Zeitspannen beliebig verändern, sowie oft ohne klare Grenzen räumlich von einem in ein anderes übergehen.

Einen Ansatz, wo keine Mechanismen der Mutation und Selektion ausgewertet werden, liefert Verschure (*Rob*, *AA* 1993). Hier werden neue, adaptive Verhaltensweisen durch Verkettung und Chunking durch i.w. konnektionistische Techniken aus vorhandenen kombiniert. So können zwar "höhere" Verhaltensweisen auf neuen Ebenen entstehen, jedoch keine qualitativ neuen auf der Ebene des vorgegebenen Basisrepertoires. Verschure demonstriert seine Ideen an einem simulierten navigierenden Roboter, dessen Performanz beim Anfahren möglichst vieler *targets* durch das Lernverfahren verbessert wird. Das Lernverfahren wird in der Simulation on-line bei fortgesetzter Aufgabenerledigung des Agenten eingesetzt. Ein sehr elaborierter Ansatz ist der von Gaudiano & Grossberg (*Bio*, *Rob*, *NN* 1991), die die Steuerung eines Roboterarmes behandeln, und Zalama et al. (*Bio*, *Rob*, *NN* 1995), der dieselben Techniken bei der Navigation eines mobilen Roboters einsetzt. In beiden Fällen geht es um das unüberwachte on-line Lernen von sensomotorischen Koordinationen (optisch geführtes Greifen bzw. Zielansteuern). Es wird ein komplexes, neurobiologisch motiviertes System aus mehreren neuronalen Netzen konstruiert, das im wesentlichen Piagets primäre Zirkulärreaktion (auf die sich die Autoren explizit berufen) realisiert. Das System kann ohne Beeinträchtiung seiner aktuellen Funktionsfähigkeit dazulernen. Der Lernvorgang bleibt jedoch auf die eine Klasse sensomotorischer Koordinationen beschränkt, für welche das dedizierte Netzwerk gebaut wird. Die Arbeit beeindruckt durch ihre Einbettung in umfassendere neurologische und funktionale Zusammenhänge und ihre "lebensnahe" Flexibilität und Robustheit.

Einen außerordentlich differenzierten Ansatz trägt Drescher (*KI* 1991) bei. Sein *schema mechanism* ist ein symbolisches Modell eines lernenden Agenten, welcher zu einer autonomen qualitativen Entwicklung sowohl von sensomotorischen Basiskonzepten als auch höheren Konzepten, die sich aus der Interaktion des Agenten mit der Umgebung ergeben, fähig ist. Da die von Drescher im Einzelfall untersuchten Konzepte aus dem Bereich der Sensomotorik stammen, wird die Arbeit an dieser Stelle vorgestellt. Der methodologische Hintergrund von Dreschers Arbeit ist die Entwicklungspsychologie Piagets (Piaget *Psych* 1947, Standardreferenz: Flavell *Psych* 1968), deren Mechanismen unter Berücksichtigung neuerer Kritiken teilweise im Detail übertragen werden. Die Erfindung neuer Konzepte wird durch eine durch und durch dynamische Auffassung von der Bedeutung von Konzepten ermöglicht. Ein Konzept hängt stets mit einer situierten Aktion des Agenten zusammen und bedeutet die Bedingungen dafür, daß die Aktion zum Erfolg führt. Die Aktionsfähigkeit und die Situationen sind (vom Agenten zunächst nicht "verstanden" oder "konzeptualisiert") vorgängig vorhanden; das zugeordnete Konzepte ist anfangs "leer" und nimmt im Laufe der Zeit durch Ausnutzung statistischer Korrelationen einen Inhalt an, welcher die Erfolgsbedingungen der Aktion repräsentiert. Das *Neue* gelangt also in das sich entwickelnde konzeptuelle System nicht durch interne Mutationen (wie bei evolutionären Ansätzen), sondern durch die zufällige Entdeckung von Regularitäten in der Agent-Umwelt-Interaktion (vgl. auch Jaeger 1994a, p. 45ff). Drescher demonstriert u.a., wie es zur Entwicklung eines Objektbegriffs aus primären sensomotorischen Reflexen kommt. Nach meiner Meinung ist die Arbeit von Drescher wegweisend. Daß sie nicht rezipiert worden ist, mag neben ihrer gedanklichen Originalität und Komplexität daran liegen, daß sie für klassische KI'ler zu wenig (nämlich überhaupt nicht) logikorientiert ist. In Dreschers Arbeit finden sich auch keine systemtheoretischen Methoden, jedenfalls nicht explizit. Der Ansatz ist hier dennoch aufgeführt, erstens, weil er sich zum Konstruktivismus bekennt, welcher hinwiederum mit Maturana und Varelas Autopoiesebegriff verbunden ist, und zweitens, weil er eine fortwährende, physikalische Agent-Umgebungs-Interaktion zur Ausgangsbasis seines Intelligenzbegriffs macht, was hinwiederum typisch für die systemorientierte Sicht auf autonome Agenten ist (vgl. Bickhard *Phil* 1993 für eine interaktionistische Deutung von Konzeptinhalten).

Betrachtet man die beiden nach meiner Meinung herausragenden Arbeiten von Drescher und Gaudiano & Grossberg, so lohnt es sich vielleicht, drei Gemeinsamkeiten herauszustellen. Erstens berufen sich beide Ansätze explizit auf Ideen von Piaget. Zweitens verwenden sie beide keinen evolutionären Mechanismus (im Sinne von Mutation und Auslese). Drittens verwirklichen sich beide im komplizierten Zusammenspiel vieler Einzelmechanismen. Sie demonstrieren, daß selbstorganisierte Adaptation eine komplexe und subtile Angelegenheit ist, die kaum durch eine Zauberformel zu erledigen sein wird.

Dalenoort (*Psych*, *Phil* 1990b) kommt in einer methodologischen Studie zum Schluß, daß Lernen als echte Selbstorganisation von "informational systems" bisher nur bei biologischen Systemen verwirklicht sei.

Kommentar

(A) Ein klassischer Anwendungsbereich systemtheoretischer Methoden ist die Analyse von Bewegungen und die Kontrolle von bewegungserzeugenden Systemen; (B) Bewegungen gehören zu einem physikalisch situierten Agenten naturgemäß hinzu; und (C) schließlich spricht manches dafür, daß physikalische Situiertheit naturgemäß zu einer wirklichen Intelligenz dazugehört. Die Folgerung für die Methodologie der KI liegt auf der Hand.

4.2.3 Konzeptuelle Informationsverarbeitung

Die Verarbeitung konzeptuellen Wissens ist die zentrale Domäne der klassischen KI. Anders als in den beiden vorangegangenen Abschnitten gibt es hier keine etablierten, systemtheoretisch orientierten Forschungstraditionen zu referieren. Aus heterogenen Quellen ergeben sich dennoch manche Hinweise darauf, wie eine systemtheoretische Sicht dieser Verarbeitungsebene aussehen und welche Vorteile sie bringen könnte.

Konzepte

In der klassischen KI werden Konzepte meistens extensional als Klassen von Referenten interpretiert (Einführung in die klassische Konzepttheorie: Smith & Medin *Ling*, *Kog* 1981). Diese Sicht hat jedoch Schwierigkeiten mit einer ganzen Reihe von Phänomenen von empirisch beim Menschen gefundenen Konzepten (Übersicht: wahrnehmungspsychologisch in Treisman *Psych* 1986, kognitionspsychologisch in Medin *Psych* 1989, Goschke & Koppelberg *Psych*, *NN* 1990). In diesem Abschnitt sollen einige der Schwierigkeiten der logik-orientierten Sichtweise angesprochen werden, die sich mit systemtheoretischen Interpretationen auflösen lassen. Dazu wird zunächst referiert, was es zur Zeit an systemtheoretisch orientierten Konzeptmodellierungen gibt.

Die wohl größte Klasse dieser alternativen Modellierungen benutzt lokalistische neuronale Netzwerke, in denen Knoten mit Konzepten, Features, syntaktischen Merkmalen, aber auch übergeordneten Kontexten gelabelt sind. Gleichgewichtszustände (d.h. Punktattraktoren) von Aktivierungen in solchen Netzwerken werden als Konzeptverwirklichungen interpretiert, wobei sich Kontexteinflüsse durch extern vorgegebene Klammerung (engl. *clamping*) von gewissen Netzknoten geltend machen. Die Tradition solcher Netzwerke läßt sich bis auf Quillians (*KI*, 1968) semantisches "Ur-" Netzwerk und Wilks' (*KI*, *Ling* 1975) spreading-activation Netzwerk zurückverfolgen. Der konnektionistische Gedanke einer sich äquilibrierenden Aktivierung kam jedoch erst bei Waltz & Pollack (*Ling* 1985) voll zum Zuge. Seine Arbeit hat bei einer ganzen Reihe von Nachfolgern Verfeinerungen erfahren. Hier seien Bookman (*KI* 1988) und Mangold-Allwinn (*Ling* 1991) genannt, wobei letztere Arbeit einen ausgezeichneten Überblick über das Gebiet enthält. Mehl (*Ling* 1992) verbindet konnektionistische und klassischere Techniken semantischer Netze einer Anwendung, bei der es um die Übergänge zwischen feinen Bedeutungsvarianten von Wörtern und deren Priming geht.

Smolenskys (*NN* 1986) *harmony theory* beruht ebenfalls auf Aktivierungsäquilibration in einem lokalistischen Netzwerk. Im Unterschied zu den vorgenannten Arbeiten leistet der Autor jedoch eine gründliche mathematische Analyse mit Mitteln der Systemtheorie und der statistischen Mechanik/Informationstheorie, indem er sein Netzwerk als dynamisches System mit einer Hamiltonfunktion (d.h. einer formalen Energiefunktion) beschreibt. Legendre et al. (*Ling*, *NN* 1990a,b) erweitern diesen Ansatz von einzelnen Konzepten auf komplexere linguistische Strukturen.

Systemtheoretisch gesprochen, werden in diesen Arbeiten Konzepte als Punktattraktoren modelliert. Die Trajektoren dieser Systeme konvergieren in der Regel rasch zu einem Punktattraktor. Dem entspricht die Verwendung der Netze für "one-shot" Klassifikationen eines einzelnen vorgegebenen Musters. Ein Input*strom* ist nicht vorgesehen.

Dieselbe Grundidee von Konzepten als Punktattraktoren findet sich auch in einigen epistemologisch orientierten Arbeiten. MacLennan (*NN*, *Kog*, *Phil* 1991) entwirft einen methodologischen Rahmen für das Verständnis von neuronalen Netzen als *continuous symbol systems*, wobei er für zentrale Eigenschaften der herkömmlichen diskreten Symbolsysteme (logik-orientierte Kalküle) korrelierende kontinuierliche, dynamische Mechanismen vorschlägt. Krause (*Psych*, *Kog* 1989) nimmt thermodynamische Prinzipien der Ordnungsbildung durch Minimierung von Entropieerzeugung zum Vorbild zur Erklärung von Ordnungsbildungen kognitiver Strukturen (u.a. Konzepte). Er geht der Frage nach, wie der "kognitive Aufwand", dessen Minimierung zur Stabilisierung kognitiver Strukturen führt, empirisch zu operationalisieren ist, und beschreibt einige Experimente. Van Leeuven (*Psych* 1990) geht u.a. von Krauses Ideen und dem gestalttheoretischen Begriff der Prägnanz (d.h. dem Grad des Hervortretens einer Gestalt aus dem Hintergrund) aus und gelangt zu einem psychologischen Modell der Wahrnehmung, das formal den oben besprochenen lokalistischen Netzwerken ähnelt. Seine Dynamik bei Klassifizierungsvorgängen gehorcht zwar einem Minimierungsprinzip, läuft jedoch nicht immer in Punktattraktoren hinein.

Auf einige Arbeiten, die in 4.2.1 und 4.2.2 zitiert wurden, muß an dieser Stelle noch einmal hingewiesen werden. Auch dort wurden sensomotorische Konzepte als Punktattraktoren modelliert, zumeist in neuronalen Netzen: bei Braitenberg (1978), Riedel et al. (1988), Buhmann & Schulten (1987), Carpenter & Grossberg (1990), und Steels (1994b). Auf die Frage, wieweit man bei sensorischen Features oder einfachen motorischen Behaviors von Konzepten sprechen kann, wird weiter unten in den Ausführungen über Attribute vs. Konzepte eingegangen.

Grenzzykel und chaotische Attraktoren sind ebenfalls als Modell für Konzepte vorgeschlagen worden. Aus vorangegangenen Abschnitten sei hier noch einmal an die vielfach verwendete Idee der Informationsrepräsentation durch synchronisierte Aktivität in Neuronenverbänden, Babloyantz' & Lourenços (1994) semistabile Grenzzykel, sowie Yao & Freemans (1990) chaotische Attraktoren erinnert.

In den bisher angeführten Arbeiten werden Konzepte als Attraktoren realisiert. Es gibt jedoch auch einige Ansätze, in denen Konzepte zwar als dynamische Phänomene, jedoch nicht (zumindest nicht offensichtlich) als Attraktoren in Erscheinung treten.

Dies gilt vor allem für die von Classifier-Systemen angebotenen Konzeptmodelle. Hier ist die Situation allerdings etwas verworren. Classifier-Systeme sind ursprünglich (Holland *Bio*, *KI* 1975) als ein allgemeines, eher biologisch motiviertes Modell adaptiver Systeme entworfen worden, aber haben schon bald durch ihren Erfinder und seine Schüler eine entschieden Betonung als Modell kognitiver Systeme erfahren. Forrest (*Inf*, *Kog* 1986, referiert nach Goldberg *Inf* 1989, 296ff) bildet KL-ONE-Netzwerke auf Classifier-Systeme ab ("CS-ONE"). Holland & Reitmann (*KI*, *Kog* 1978, referiert nach Goldberg *Inf* 1989, 265ff) entwerfen ein lernfähiges, durch Motivationsvariable moduliertes Wahrnehmungs-Klassifizierungs-Entscheidungs-Aktionssystem (CS-1 für "Cognitive System One"), dessen Leistungen sie in einer simulierten Labyrinth-Futtersuchaufgabe analysieren. Eine vereinfachte Version wird bei Holland (*KI*, *Inf* 1986) vorgestellt. Im Gegensatz zu der möglicherweise in der KI vorherrschenden Meinung sind Classifier-Systeme also von Hause aus nicht (nur) für langfristige, evolutionäre Adaptationsmodelle intendiert, sondern (auch) für die on-line Kontrolle von situierten Agenten.

Holland (*KI*, *Inf* 1986) spricht allgemein von Konzeptrepräsentationen durch Mengen ko-aktivierbarer Classifier. Er beschreibt eine ganze Reihe von Kopplungsmechanismen zwischen Classifiern, ohne eine einheitliche Definition zu liefern. In allen eben erwähnten Arbeiten werden Konzepte jedenfalls in einer Art verteilter Repräsentation durch Mengen von Classifiern modelliert. Nach meiner Interpretation scheint es mindestens drei Grundtypen zu geben. Der einfachste sind Mengen von Classifiern, die wegen Übereinstimmungen in ihrem Bedingungsteil gleichzeitig aktiviert werden können. In klassischer Sichtweise entsprechen sie am ehesten einer Konzeptmodellierung durch eine Menge von Merkmalen. Sodann gibt es zyklisch geschlossene Ketten von Classifiern. Sie sind in der kurzfristigen Dynamik des Basis-Arbeitszyklus' (Holland 1986) selbststabilisierend und können wohl als Attraktoren aufgefaßt werden (wie bei Patel & Schnepf *KI* 1991). Ein dritter Typ von Konzeptrepräsentationen wäre offene Ketten von Classifiern. Sie wären im Sinne dynamischer Systeme als Transiente aufzufassen.

Die dynamischen Eigenschaften solcher Repräsentationen sind schwer beherrschbar und noch lange nicht aufgeklärt (Forrest & Miller *Inf, Math* 1990). Es gilt mindestens drei Zeitskalen zu beachten, nämlich die kurzfristige Dynamik des *basic execution cycle,* welche den Arbeitsmodus eines Classifier-Systems ohne Berücksichtigung von Lernvorgängen ausmacht; die relativ kurzfristige Adaptation bzw. Optimierung des Systemverhaltens durch die *bucket-brigade,* einen *apportionment-of-credit*-Algorithmus, welcher für die aktuelle Aufgabe besonders geeignete, vorhandene Classifier verstärkt; und die längerfristige Adaptation durch genetische Algorithmen, welche durch Mutationsvorgänge zu qualitativen Neuerungen in der Classifierpopulation führen kann (Holland *KI*, *Inf* 1986).

Eine Einzelstellung nimmt das Copycat-System von Hofstadter & Mitchell (*KI*, *Kog* 1993) ein. Sie verwenden eine ausgetüftelte Kombination verschiedener Techniken und Module, um die Fähigkeit zu Analogieschlüssen zu modellieren. Die Entdeckung einer Analogie zwischen zwei Reizmustern geschieht in einem kombiniert konnektionistischen, symbolverarbeitenden und evolutionären Prozeß, der nach einer u.U. sehr komplexen und nichtdeterministischen Transiente sich fast immer in einem Vorschlag für eine Analogie stabilisiert. Dabei können überraschende, qualitativ vom Programmierer nicht "vor-gedachte" Entdeckungen gemacht werden. Copycat weist Besonderheiten auf, die es mit keinem anderen mir bekannten System teilt. In einem neuronalen Netzwerk, das Langzeitinformation enthält, entsprechen Links Knoten, und verändern ihre Leitfähigkeit (als Links) in Abhängigkeit von ihrer Aktivierung (als Knoten). Der Verarbeitungsprozeß insgesamt steht unter dem Einfluß einer *computational temperature*, die in Abhängigkeit von der "Kohärenz" des gerade erreichten Zwischenergebnisses steigen oder fallen kann, was ein Novum gegenüber dem häufiger zu findenden *simulated annealing* ist. Konzepte werden in Copycat in mehreren Sorten und mit verschiedenen Techniken realisiert, von gelabelten Knoten im Langzeitgedächtnis bis zu Populationen sehr einfacher informationstragender "Gene" [Metapher von Hofstadter und Mitchell], sogenannten *codelets*. In diesem letzteren Sinn sind Konzepte systemtheoretisch als Punktattraktoren deutbar. Allerdings ist der Äquilibrierungsprozeß so komplex und instabil, daß eine formale Attraktorrekonstruktion sicher nicht einfach ist. Die Verfasser sind sehr liberal in ihrem Sprachgebrauch; einen durchgängig eingehaltenen Begriff von "Konzept" sucht man vergeblich. Zu Copycat existiert keine einheitliche mathematische Formalisierung oder Analyse. Sie wäre wegen der Heterogenität und Komplexität des Systems auch kaum zu leisten.

Ebenfalls ein Einzelansatz zur Konzeptmodellierung ist der schon in 4.2.2 vorgestellte von Drescher (*KI* 1991). Konzeptinhalte sind dynamische Regularitäten von situationsgebundenen Aktionen. Konzepte differenzieren sich unter wechselseitiger Konkurrenz aus (die Konkurrenz entsteht implizit durch die Abhängigkeit der Konzeptweiterentwicklung von Aktionsausführungen, die miteinander konkurrieren), und es entwickelt sich eine große Population von Konzepten (viele Tausend in ersten Simulationen). Drescher versucht keine mathematische Analyse des Systemverhaltens; Methoden der Populationsdynamik bieten sich eventuell an.

Ich habe mich selbst ausgiebig mit einer systemtheoretischen Modellierung von Konzepten beschäftigt (Jaeger *KI*, *Math* 1991, 1992). Inzwischen ist dabei eine Klasse von diskreten dynamischen Systemen ("dynamische Symbolsysteme") herausgekommen, die als Modell selbstorganisierender Informationsverarbeitung in situierten Agenten gedacht ist (Jaeger 1994a,d). Dies Modell beschreibt miteinander durch top-down und bottom-up-Wechselwirkungen gekoppelte, formal gleich geartete Verarbeitungsebenen, die von der sensomotorischen Peripherie bis zu einem Konzepte verarbeitenden Zentrum reichen und insgesamt eine sog. *associety* bilden. Jede Verarbeitungsebene kann in erster Näherung als eine räumlich strukturierte Population von diskreten Verarbeitungseinheiten ("dynamische Symbole") vorgestellt werden, die sich aufgrund lokaler Gesetzmäßigkeiten dynamisch verändert. Dabei entstehen in jeder Verarbeitungsebene durch Rückkopplungseffekte semistabile Muster, sog. "Resonanzen", denen in der jeweils höheren Ebene ein einzelnes, "gröberes" dynamisches Symbol entsprechen kann. Konzepte werden durch solche Resonanzen modelliert, samt den entsprechenden gröberen dynamischen Symbolen in der nächsthöheren Ebene sowie zunehmend "feinkörnigeren" Entsprechungen in den tieferen Ebenen. Innerhalb einer Ebene ist so eine Resonanz als Attraktor deutbar; ein evtl. zugeordnetes grobkörnigeres dynamisches Symbol auf der nächsthöheren Ebene ist dagegen entweder als ein räumlicher Übergang, eine zeitliche Transiente, oder einer Kombination daraus zu sehen. Diesem Attraktor/Transienten-Dualismus entsprechen unterschiedliche Zeitmaßstäbe in verschiedenen Ebenen. Der Ansatz ist formal einheitlicher und einfacher als die anderen hier vorgestellten Einzelentwürfe, die im übrigen in mancher Hinsicht vergleichbar sind (Diskussion in Jaeger 1994a p.146). Er enthält jedoch gegenwärtig keine Lernfähigkeit. Ebenfalls steht eine Analyse dynamischer Symbolsysteme mit formalen systemtheoretischen Techniken noch aus.

Zusammenfassend läßt sich zu den bisher in der Literatur zu findenden, systemtheoretisch deutbaren Ansätzen zur Konzeptmodellierung folgendes sagen:

- Zumeist werden Konzepte als Punktattraktoren aufgefaßt, und hierbei meistens in lokalistischen konnektionistischen Netzwerken mit einer Äquilibrationsdynamik.
- Seltener werden chaotische Attraktoren und Grenzzykel vorgeschlagen.
- Einige Autoren entwerfen relativ komplexe, diskrete Agentenmodelle, in denen Konzepte eine reichhaltigere dynamische Phänomenologie aufweisen, die nicht mehr in ein einfaches Attraktormodell gefaßt werden kann. Die mathematische Analyse solcher Architekturen als dynamische Systeme hat aber noch kaum begonnen.

Nun sollen einige mit Konzepten verbundene Phänomene betrachtet werden, mit denen die klassische, logik-orientierte Konzeptmodellierung Schwierigkeiten hat, die sich jedoch systemtheoretisch gut fassen lassen.

Empirisch beim Menschen beobachtete Konzepte sind auf verschiedenen Maßstäben **zeitvariabel**. Einige klassische Befunde seien hier erwähnt. Lakoff (*Ling* 1987) beschreibt, wie in einem Zeitraum von Generationen Konzepte eine *radiale* Bedeutungsstruktur erhalten, in einem Vorgang, der grob mit der biologischen Entstehung von Artenstammbäumen verglichen werden kann. Bartlett (*Psych* 1932) untersucht auf der Zeitskala von Monaten und Jahren Erinnerungsprozesse und findet, daß Konzepte sich von den Vpn unbemerkt in Richtung einer größeren subjektiven Kohärenz organisieren. Barsalou (*Ling*, *Kog* 1987) beschreibt die Gradiertheit (engl. *gradedness*) von Konzepten, d.h. die Existenz von sowohl zentralen als auch peripheren Vertretern von Kategorien (z.B. wird *robin* als ein typischer, *ostrich* dagegen als ein untypischer Vogel bewertet). Typikalitätsratings ändern sich intrapersonal im Lauf von Wochen und unter dem Einfluß von Kontextinformation. Auf der sehr kurzen Zeitskala von einigen Zehntelsekunden finden schließlich Seidenberg et al. (*Psych* 1982), daß ein Konzept beim Aufruf aus dem Gedächtnis in einem Stück geladen, sondern aus verschiedenen

Informationsbausteinen und mit verschiedenen Mechanismen zusammengesetzt wird, wobei wieder der Kontext anderer aktualisierter Konzepte mitbestimmend ist.

Befunde dieser Art lassen sich nicht gut mit einer logik-basierten Konzeptmodellierung vereinbaren, da logische bzw. modelltheoretische Bestimmungen inhärent zeitlos sind (Diskussion in Jaeger *KI*, *Math* 1992, 1994a, p.134ff). Dagegen nehmen systemtheoretische Modellbildungen gerade von der Zeitlichkeit der betrachteten Systeme ihren Ausgang. Sie sind daher der Ansatz der Wahl, wenn die Zeitlichkeit von Konzepten ernst genommen wird. Dieser Aspekt wird van Gelder & Port (*Kog* 1994b) als zwingender Grund für die Verwendung dynamischer Systeme in der Kognitionswissenschaft herausgestellt.

Wie bereits in 4.2.1 bemerkt wurde, liefert die formale Modellierung eines konzeptuellen Systems als dynamisches System keineswegs sogleich ein Modell der zeitlichen Phänomene. Diese müssen erst im einzelnen definiert und modelliert werden, wozu freilich dynamische Systeme dann die besten Voraussetzungen mitbringen. Die Phänomene reichen von kurzfristigen Äquilibrierungs-, Synchronisierungs- und Oszillationsvorgängen über mittelfristiges Lernen und Adaptation bis zu Generationen überdeckenden, evolutionären Prozessen, wozu es verschiedene systemtheoretische Mechanismen gibt. Eine einheitliche Theorie der phänomenalen Zeitlichkeit in dynamischen Systemen ist nicht in Sicht und (nach allem, was uns vor allem die Evolutionstheorien über den hier waltenden Erscheinungsreichtum ahnen lassen) wohl auch nicht möglich. Komplexere Ansätze zur dynamischen Modellierung kognitiver Systeme (Carpenter & Grossberg 1990 und Gaudiano & Grossberg 1991, Yao & Freeman 1991, Drescher 1991, Holland & Reitman 1978, Hofstaedter & Mitchell 1993, Jaeger 1994a) enthalten mehrere zeitliche Mechanismen auf verschiedenen Skalen und demonstrieren die Beschränktheit der heute möglichen theoretischen Durchdringung. Es ist aber sicher ein Verdienst systemtheoretisch orientierter Denkweisen, die zeitliche Phänomenologie von Konzepten überhaupt sichtbar gemacht zu haben.

Eine Art von **Kontextabhängigkeit** von Konzepten äußert sich darin, daß beim Aufruf eines Konzeptes in Abhängigkeit von gleichzeitig oder vorher aktualisierten Konzepten verschiedene Merkmale salient werden. In der Kognitionswissenschaft und der Psycholinguistik gibt es hierzu viele empirische Untersuchungen (Sammlungen und Diskussionen in Mangold-Allwinn *Ling* 1991, Barsalou *Kog* 1987, 1989, wahrnehmungspsychologisch orientiert in Treisman *Psych* 1986 p. 35-44, Richards *Psych* 1988, Goschke & Koppelberg *Ling, NN* 1990). Die einfachste, jedoch unbefriedigende Methode, wie man in der logik-orientierten KI mit der Kontextvariabilität von Konzepten fertigwerden kann, sind probabilistische Attributzuweisungen (Smith & Medin *Kog* 1981). Um das Problem bei den Wurzeln zu packen, muß man aber wohl die gewohnte Prädikatenlogik aufgeben und zum Beispiel Situationssemantiken oder nichtmonotone Logiken einsetzen. Dadurch handelt man sich allerdings im allgemeinen rechnerische Intraktabilität und (nach meiner Ansicht) wenig plausible und intransparente Modelltheorien ein. In technischen Anwendungen hilft man sich, indem man verschiedene zu erwartende Kontexte explizit als Frames o.ä. modelliert und die verschiedenen Ausprägungen eines Konzepts, das in mehreren solcher Kontexte vorkommt, durch verschiedene Defaultwerte in diesen Frames spezifiziert. Dies ist ein pragmatischer "Hack" ohne theoretische Basis. Beispiele zu diesen Strategien finden sich in Brézillon (*KI* 1993).

Systemtheoretisch lassen sich Kontexteinflüsse auf mehrere Weisen modellieren. Die gängigste Art ist in lokalistischen neuronalen Netzwerken zur "one-shot"-Klassifikation verwirklicht, in denen Konzepte Punktattraktoren sind (s.o.). Hier tauchen Kontexteinflüsse als Kontrollparameter auf (in der Form von festgehaltenen [engl. *clamped*] Inputknoten-Aktivierungen). In einem gewissen Bereich führt die Variation dieser Kontrollparameter zu stetigen Verschiebungen der Attraktoren; bei Abweichungen über kritische Werte hinaus kommt es zu Bifurkationen. Dem entsprechen eine kontextabhängige, stetige Verschiebung von Konzeptinterpretationen und plötzliche Umdeutungsereignisse. Eine solche qualitative

Differenz ist zum Beispiel die Interpretation von *buck* als Hirsch oder Dollarnote bei Waltz & Pollack (*Ling* 1985).

Das Prinzip "Kontext durch Kontrollparameter" läßt sich natürlich auch bei komplexeren Attraktormodellen für Konzepte verwenden. Eine besonders interessante Variante könnte die von Babloyantz und Laurenço (*Bio* 1994) vorgeschlagene Konzeptmodellierung durch reizstabilisierte Zykel in einem chaotischen Grundzustands-Attraktor abgeben (die Autoren haben dabei wahrnehmungspsychologische Kategorien, nicht kognitive Konzepte im Sinn). Je nach Reizzusammensetzung können verschiedene Zykel simultan stabilisiert werden, d.h. es findet eine diskretisierende, kompositionale Klassifikation statt. Ein Konzeptaufruf ist hier also eine spontane, Input-getriebene Zusammenstellung von etwas, das man Features nennen könnte. Diese ist kontextsensitiv, insofern der Input Kontextinformation enthält.

Der Kontext eines aufgerufenen Konzepts kann in nicht-konzeptueller Information bestehen, z.B. in sensorischem Input oder motivationalen Einstellungen. Dann ist der eben angeführte Mechanismus "Kontext durch Kontrollparameter" plausibel, falls die Kontextinformation relativ zeitstabil ist. Nun ist das wohl nicht typischerweise der Fall. Erstens ist der konzeptmodifizierende Kontext oft selbst schon konzeptueller Natur — etwa wenn er in schon konzeptualisiertem Input oder in ko-aktivierten Konzepten besteht. Zweitens kann sich der Kontext durchaus auf demselben Zeitmaßstab verändern wie die beeinflussten Konzepte. In diesem Fall sind Kontrollparameter nicht mehr als formales Pendant von Kontexteinflüssen geeignet.

Statt dessen sind hier Modellbildungen plausibler, in denen mehrere Konzepte interagieren und wechselseitig füreinander den Kontext bilden. Diese Situation besteht bei Konzeptrepräsentationen, die aus Populationen von Informationsträgern bestehen. Beispiele sind Cluster von Classifiern oder von *codelets* in Copycat (s.o.). In solchen Populationen kann simultan das Material für mehrere Konzepte vorliegen. Bei entsprechenden Konkurrenzmechanismen (die in diesen beiden Beispielen gegeben sind) kommt es zu Interaktionen zwischen Konzepten, oder, anders ausgedrückt, zu wechselseitigigen Kontextbildungen. Wie bereits oben angedeutet, ist die mathematische Analyse ist hier noch nicht sehr weit.

Eine systemtheoretische Rekonstruktion von Konzept-Konzept-Wechselwirkungen wird am Ende nicht um die Modellierung von "räumlichen" und "kinematischen" Beziehungen zwischen Konzepten umhinkommen. Dabei ist nicht unbedingt an metrische Räume gedacht; auch schwächere topologische Beziehungen zwischen einzelnen aktiven Konzepten könnten Bedingungen wechselseitiger Einflüsse modellieren.

In biologischen neuronalen Netzen ist die räumliche Struktur funktional. Manchmal lassen sich bestimmte Konzepte bestimmten, kleinräumigen Arealen zuordnen, z.B. bei der bekannten somatotopischen Projektion (vgl. Schmidt R.F. *Bio* 1980, 71ff) oder bei Verlusten von Benennungen bestimmter Kategorien (z.B. von Frucht- und Gemüsenamen, berichtet von Hart et al. *Psych*, *Bio* 1985). Im visuellen Kortex gibt es eine kleinräumige Feature-Spezialisierung von Neuronenverbänden, deren Genese mit *self-organizing feature maps* in künstlichen Netzwerken rekonstruiert werden kann (Obermeyer et al. *NN*, *Bio* 1990). In einem künstlichen neuronalen Netz, das eine Komponente in einem Neurokontroller für visuell geführte Greifbewegungen ist, macht sich Kuperstein (*Rob*, *NN* 1991) dasselbe Prinzip zunutze. Neben solchen Befunden einer räumliche Lokalisierung von Konzepten oder verwandten Phänomenen gibt es jedoch bekanntermaßen auch Befunde, die auf eine großräumig verteilte, "holistische" Repräsentation deuten. Insgesamt kann man wohl davon ausgehen, daß es in biologischen Gehirnen mehrere räumliche Ordnungsprinzipien gibt, die auf noch kaum geklärte Weise nach Hirnareal, Konzeptsorte u.a. variieren. Außerdem sind Wechselwirkungen räumlicher Organisation mit zeitlichen Phänomenen nach meiner Meinung wichtig, aber noch nicht systematisch untersucht. Erste Ansätze sind z.B. die räumliche Speicherung zeitlicher Sequenzen in einem Kohonen-Netz mittels *leaky-integrator*-Neuronen (Campbell & Taylor 1993) und das schon in 4.2.1 erwähnte Netzwerk von Matsuga & Yuille (*NN*, 1994).

Zeitveränderliche räumliche Beziehungen zwischen Konzepten findet man bei den bereits eingeführten Systemen von Hofstadter & Mitchells (1993) Copycat und bei meinen (Jaeger 1994a) Associeties. In Copycat's Langzeitspeicher sind Konzepte als Knoten in einem Graph mit definierten Längen repräsentiert. Diese Längen reflektieren die assoziative Nähe der Konzepte; die Nachbarn eines Konzeptes bilden eine probabilistische "Wolke" um dies Konzept, welche mit dem Konzept"inhalt" selbst identifiziert wird. Durch Konzeptaktivierungen werden Veränderungen der Kantenlängen und damit der aktuellen Konzeptinhalte induziert, was hinwiederum Auswirkungen auf die Aktivierungen hat. In Associeties sind die einzelnen Verarbeitungsebenen formal gerichtete, kantengelabelte Graphen. Konzepte sind, grob gesagt, zyklische Teilstrukturen solcher Graphen. Durch die Graphstruktur werden Wechselwirkungsmöglichkeiten zwischen Konzepten definiert. Die Wechselwirkungen können mehrere Effekte zeitigen (z.B. Anreicherung, Austausch, Verschmelzung, Duplizierung, Löschung von Bausteinen eines Konzepts). Die Graphstruktur selbst verändert sich im Laufe der Interaktionen, so daß es u.a. zu relativen "Bewegungen" der Konzepte kommen kann. Auch hier ist eine mathematische Analyse der raumzeitlichen Aspekte der Dynamik noch nicht geleistet.

Generell ist zu bemerken, daß raumzeitliche Prozesse ein zwar zentrales, aber leider auch besonders schwieriges systemtheoretisches Thema sind. In der Synergetik werden die grundlegenden Definitionen eingeführt und einige sehr einfache Systemklassen hinsichtlich Bifurkationen analysiert. Dazu konstatiert Haken (*Phys* 1983, p.272): *"So far not very much has been done in this field, but I think that such a treatment will be indispensable,* [...]*"*. Neben kontinuierlichen Systemen werden auch die handlichen Zellularräume (engl. *cellular automata*) als diskrete Modelle räumlicher dynamischer Systeme in den Naturwissenschaften gerne verwendet (Sammlung von theoretischen Analysen und Anwendungsbeispielen in Wolfram *Math* 1986). Dabei können selbstorganisierenden Musterbildungen bequem simuliert werden. Ein Beispiel sind Diffusions-Reaktionsprozesse in erregbaren Medien (Gerhardt & Schuster *Math* 1989), eine Prozeßklasse, die auch für die Beschreibung von Erregungsmustern in neuronalen Netzen von Interesse ist. Auch bei Zellularräumen steckt die mathematische Analyse von raumzeitlichen Phänomenen noch in den Anfängen (Packard & Wolfram *Math* 1985).

Zusammenfassend läßt sich zum Thema Kontextsensibilität festhalten:

- Kontextsensitivität von Konzepten ist kein scharf definierbares Phänomen, sondern umschreibt — ebenso wie übrigens der Konzeptbegriff selbst — eine ganze Gruppe von empirischen Erscheinungen.
- Kontextsensitivität ist ein dynamisches Phänomen, das sich in einer zeitlichen Variabilität von Konzepten äußert. Zu seiner Behandlung sind systemtheoretische Ansätze daher eine natürliche Wahl.
- Verbreitet, aber zu einfach ist die Modellierung von Kontextsensitivität mittels lokalistischer neuronaler Netze.
- Adäquatere Modellierungen müssen wohl in der einen oder anderen Form räumliche oder topologische Nachbarschaftsbeziehungen zwischen Konzepten berücksichtigen. Räumliche dynamische Systeme sind jedoch mit den heute vorhandenen Methoden nicht befriedigend mathematisch zu analysieren.

Zyklische, iterierte und selbstreferentielle Beziehungen zwischen Konzepten sind verbreitet und faszinierend. Das Buch, das wohl mehr als jedes andere junge Studenten zur KI gebracht hat, Douglas Hofstadters "Gödel, Escher, Bach" (*KI*, *Kog* 1979), ist eine Orgie der Selbstbezüglichkeit. Versucht man jedoch, konzeptuelle Informationsverarbeitung in logikorientierten Formalismen wie etwa KL-ONE-artigen Sprachen zu implementieren, so befindet man sich rasch in Schwierigkeiten. Sie gründen praktisch in algorithmischen Problemen mit Zykeln, und theoretisch darin, daß man für die modelltheoretische Semantik eine fundierte

Mengenlehre verwendet (Nebel *KI* 1990, 1991, weitere Diskussion in Jaeger *KI*, *Math* 1994a, aber: nichtfundierte Mengenlehre zur Modellierung von mutual beliefs bei Barwise *Math*, *Kog* 1989, allgemeineres Plädoyer für nichtfundierte Mengenlehre bei Barwise & Moss *Math* 1991).

Demgegenüber sind Selbstbezüglichkeit und Zyklizität in Formalismen dynamischer Systeme von vorneherein eingebaut. In Systemen von Differentialgleichungen $d\mathbf{x}/dt = \mathbf{f}(\mathbf{x})$ wird die Veränderung der Observablen $\mathbf{x}$ (ein Vektor mehrerer Variable!) auf die Werte derselben Variablen zurückgeführt, Diffeomorphismen generieren eine Systemdynamik durch iterierte Anwendung einer Funktion auf das jeweils vorher erzielte Auswertungsergebnis, in diskreten Systemen sind Zustandszykel völlig normal, das klassischste aller dynamischen Systeme, das Pendel, ist periodisch. Heinz von Foerster (*Bio*, *Math*, *und alles andere* 1984a,b) verwendet in seinen schönen, amüsanten Einführungen in die Systemtheorie, die er vor den verschiedensten Auditorien gibt, Paradoxien der Selbstbezüglichkeit und ihre systemtheoretische Auflösung als Aufhängungspunkt.

(Schein-)Probleme der Zyklizität und der Selbstbezüglichkeit werden in systemtheoretischer Sicht durch eine Würdigung der Rolle der Zeit aufgelöst. Bei Differentialgleichungen, Diffeomorphismen, diskreten Systemen fallen die oben angesprochene Phänomene der Selbstbezüglichkeit mit dem jeweiligen Mechanismus der zeitlichen Dynamik zusammen. Dynamische Systeme entwickeln sich aus sich selbst weiter - das ist nicht weiter enigmatisch, sondern eine knappe Charakterisierung der Wirkungsweise von Differentialgleichungen etc.

Zyklizität und Selbstbezüglichkeit sind in der KI, der Linguistik und der Kognitionswissenschaft keine prominenten Themen. Da systemtheoretische Ansätze in diesen Disziplinen erst am Anfang stehen, nimmt es nicht wunder, daß es noch kaum formale Arbeiten zur systemtheoretischen Modellierung solcher Phänomene gibt. Man könnte auf Piaget hinweisen, dessen Theorie der Kreisreaktionen und der wechselseitigen Abhängigkeit verschiedener Sinnsesmodalitäten iterative, selbstbezügliche und regulatorische Aspekte beim Erwerb des Objektbegriffs herausstellt (Piaget *Psych* 1947, vgl. auch Drescher *KI* 1991). Der mathematische Hintergrund ist hier jedoch wenig ausgearbeitet und nicht im engeren Sinn systemtheoretisch.

Dieselbe Grundhaltung findet sich in verschiedenen Ausprägungen auch in einer Reihe von Arbeiten aus verschiedenen anderen Gebieten wieder, wo *interaktionistische* Konzeptsemantiken verfolgt werden. Dort wird der repräsentationale Gehalt von Konzepten nicht in modelltheoretischer Manier in einer ideellen Referenzbeziehung zwischen Begriff (bzw. dessen Symbol) und externem Denotat gesehen. Vielmehr resultiert der repräsentationale Gehalt eines Konzepts aus Invarianten in der Interaktionsgeschichte eines Agenten mit äußeren Objekten. "Konzepte" und "repräsentierte Objekte" sind wechselseitig aufeinander angewiesen; beide zusammen *sind* ein einziges dynamisches Interaktionsmuster. Versionen dieser Auffassung sind weit verbreitet, z.B. in der der behavior-orientierten Robotik (Verschure *AA*, *Rob* 1993, Scheier & Pfeifer *AA*, *Rob* 1995), in der Biokybernetik und konnektionistischen Robotik (vgl. die schon vorgestellten Arbeiten von Gaudiano & Grossberg *Bio*, *Rob*, *NN* 1991 und Zalama et al. *Bio*, *Rob*, *NN* 1995 zum Erlernen sensorimotorischer Koordinationen), in der Theorie autonomer Agenten und in der Epistemologie der Kognitionswissenschaft (sehr profund: Bickhart *Phil* 1993), in der KI (z.B. in der schon behandelten Arbeit von Drescher *KI* 1991), und natürlich in der Entwicklungspsychologie Piaget'scher Prägung.

Im radikalen Konstruktivismus (Maturana & Varela *Bio*, *Phil* 1984, Schmidt *Phil* 1987a) werden die Dinge ähnlich gesehen, jedoch wird die Rolle der Agent-internen Dynamik der Interaktion zwischen internen Prozessen stärker betont als die Agent-Umwelt-Interaktion. Eine weitere Variante zu einer interaktionistischen Interpretation von Konzeptinhalten hebt hervor, daß Symbole Bedeutung durch den Wert gewinnen, den ihr Gebrauch für eine Steigerung der Überlebensfähigkeit eines Agenten erbringt. In diesem Sinn untersucht MacLennan (*AA* 1990) die Emergenz von Kommunikation in simulierten Systemen von Einfachst-Agenten in einer Einfachst-Umgebung. Ähnlich motiviert sind Arbeiten von McFarland (*Bio* 1994), der die Emergenz von Kommunikationsmustern in Gruppen behavior-basierter Roboter unter

Bedingungen von Energieknappheit und dadurch implizit erzwungener Kooperation untersucht. Hier spielen "motivationale" Parameter innerhalb der Agenten die Rolle von Konzepten [meine Interpretation]; ihre Bedeutung gewinnen sie durch ihre funktionale Einbindung in die interen Kontrolldynamik.

Ohne den Anspruch, etwas zu den tieferen Fragen nach der Natur repräsentationalen Gehaltes beitragen zu wollen, diskutiere ich in Jaeger (1994a) vor dem Hintergrund der dynamischen Symbolsysteme zwei mehr oberflächliche Phänomene der Zyklizität. Dabei geht es erstens um zyklische Verweisstrukturen innerhalb einer "Resonanz", d.h. einer gestaltartigen, aus mehreren Konzepten zusammengesetzten mentalen Repräsentationsstruktur. Solche zyklische Repräsentationen können zu periodischen Verhaltensäußerungen führen. Zweitens geht es um ein formales Modell der zyklischen Verweisstruktur etwa in einer Enzyklopädie, in der ja Konzepte zunächst durch andere Konzepte, letztlich aber zyklisch durch sich selbst bzw. das Gesamtgefüge erklärt werden.

In der kognitionswissenschaftlichen, linguistischen und der KI-Literatur werden in verstreuten Bemerkungen Konzepten gerne **Gestalteigenschaften** zugesprochen. Konzepte erscheinen nicht als zufällige Anhäufungen von Merkmalen, sondern sie sind "geschlossen", "kohärent" oder eben "gestalthaft". Nun ist der Gestaltbegriff leider nicht präzise operationalisiert. Er ist seiner Herkunft nach philosophischer Natur (von Ehrenfels *Psych Phil* 1890, historischer Abriß der philosophischen Tradition in Wellek *Phil* 1972). In der klassischen, vom Dreigespann Köhler/Wertheimer/Koffka gegründeten Gestaltpsychologie (Einführung in Nyman *Psych* 1966) sind zwar viele empirische Untersuchungen zur Wahrnehmung von Gestalten gemacht worden. Zentrale Begriffe wie *Abgesetztheit*, *Geschlossenheit*, *Gestalthöhe*/*-reichtum*, *Gestaltgüte*/*-festigkeit*, *Gerichtetheit*, *Gegliedertheit*, *Stabilität* und die sog. *Gestaltgesetze* sind jedoch weder formal, noch unabhängig von bestimmten Aufgabenkontexten, noch unabhängig von Introspektionsmethoden definiert worden. So ist es kein Wunder, daß "Gestalthaftigkeit" immer noch eher eine Metapher als eine Erklärung ist. Daß diese Metapher so regelmäßig verwendet wird, zeigt aber die Wichtigkeit dieses Phänomens an.

Es gibt viele lockere Querverbindungen zwischen Systemtheorie und Gestaltpsychologie. Kaum ein Systemtheoretiker der alten Schule läßt einen Hinweis in dieser Richtung aus. Von Bertalanffy (*Bio* 1968 p.208) sieht seine allgemeine Systemtheorie historisch als Parallelentwicklung zur deutschen Gestaltpsychologie. Wiener (*Ing*, *Math*, *Bio*, *Kog* 1961) widmet der visuellen Gestalterkennung ein eigenes Kapitel, wobei interessanterweise auch eine von McCulloch entwickelte frühe Version des Perceptrons besprochen wird. Wiener gibt an, daß die eher neurophysiologisch ausgerichtete Kybernetik notwendig durch die psychologische Perspektive ergänzt werden muß. So kam es auch in der Entstehungszeit der Kybernetik zu einer Zusammenarbeit mit dem amerikanischen Gestaltpsychologen Kurt Lewin. Eigen & Winkler (*Bio* 1975, Kap. 6) stellen in einem lockeren Übersichtskapitel über verschiedene chemische und biologische Strukturbildungen wiederholt Querverbindungen zum Gestaltbegriff her.

Umgekehrt wird aus der Psychologie heraus gelegentlich eine Verbindung von der Gestaltpsychologie zu systemtheoretischen Gedanken hergestellt. Krause (*Psych* 1989) gibt an, daß sein systemtheoretischer Entwurf einer Denkpsychologie dort anknüpft, wo die klassische Gestaltpsychologie aus methodologischen Gründen scheiterte. Stadler & Kruse (*Psych* 1986) zeigen methodologische und historische Verwandtschaften zwischen der psychologischen Gestalttheorie und dem systemtheoretischen Konstruktivismus (Maturana/Varela und Nachfolger) auf. In Stadler & Kruse (*Psych* 1990) beschreiben dieselben Autoren in ähnlicher Weise Similaritäten zwischen der Gestalttheorie und der Synergetik.

In allen aufgeführten Arbeiten bleiben die Querbezüge zwischen Gestalttheorie und naturwissenschaftlicher Systemtheorie oberflächlich. Hier wird meiner Meinung nach eine doppelte Chance vertan. Erstens wäre die in der Gestaltpsychologie erschlossene, reiche Phänomenologie von Perzepten und Konzepten für ein systemtheoretisches Konzeptmodell nutzbar zu machen. Die in systemtheoretischen (zumeist konnektionistischen) Ansätzen der Kognitionswissenschaft und der KI verbreitetete, aber zu einfache Modellierung von Konzepten durch Attraktoren (vgl. Abschnitt 4.3) könnte durch die Aufnahme des in der Gestaltpsychologie gesammelten Materials differenziert und angereichert werden. Umgekehrt könnten zweitens systemtheoretische formale Modelle zu der sehr wünschenswerten Präzisierung des Gestaltbegriffs in der Psychologie führen. Solche Modelle könnten insbesondere den Prozeßcharakter von Gestaltwahrnehmungen erschließen. Die übliche Gestaltpsychologie beschäftigt sich gewöhnlich nur mit den Endprodukten von Wahrnehmungs- und Generierungsprozessen, eben den fertigen Gestalten.

Einen ersten Schritt zu einer Formalisierung und Dynamisierung gestalttheoretischer Konzepte unternehme ich in Jaeger (*Math*, *KI*, *Kog* 1994b) vor dem Hintergrund des speziellen Systemtyps der dynamischen Symbolsysteme. Allgemein gesprochen ist es naheliegend, die beiden grundlegenden Gestaltkritierien der Übersummenhaftigkeit und der Transponierbarkeit durch die systemtheoretischen Begriffe der nichtlinearen Kopplung von Teilsystemen bzw. der strukturellen Stabilität zu präzisieren.

Dynamische Systeme gestatten es, die **Übergänge zwischen verschiedenen Organisationsebenen** zu erklären. Unter "Organisationsebenen" verstehe ich hier eine Schichtung von einer sensomotorischen Basisebene zu einer rationalen, konzeptuellen Verarbeitungsschicht, wie sie in vielen Variationen in der Literatur anzutreffen ist.

In der logikorientierten KI werden Übergänge zwischen verschiedenen Organisationsebenen zwar *beschrieben*, aber nicht *erklärt*. So kann man z.B. durch die Angabe entsprechender Axiome beschreiben, welche Konzepte (die einer "höheren" Organisationsebene angehören) aus welchen Attributen (einer "niedrigeren" Ebene) zusammengesetzt sind. Die Frage, woher diese Axiome kommen, kann jedoch im Rahmen der logikorientierten KI nicht gestellt, geschweige denn beantwortet werden.

Im Unterschied dazu ist die Entwicklung systemtheoretischer Techniken in den Naturwissenschaften oft gerade durch die Frage motiviert, wie die Entstehung höherer Organisationsebenen zu erklären ist (allgemeine Bemerkungen: von Bertalanffy *Bio* 1968, 27ff, Haken *Phys* 1983, p. 18ff; Klassiker: Eigen & Schuster *Bio* 1977; Ansatz einer allgemeinen Theorie der Emergenz: Baas *Math* 1994). Inzwischen sind viele mathematische Techniken verfügbar, welche auf verschiedenen Aspekten und Typen dieses Problems anwendbar sind.

Eine Gruppe solcher Techniken betrifft Systeme gekoppelter Differentialgleichungen. In solchen Systemen kann mitunter eine Aufteilung der beteiligten Zustandsvariablen nach einer höheren und einer niedrigeren Organisationsebene vorgenommen werden. Der einfachste Fall ist vielleicht der, wo wenige "höhere" Variable in ihrem Verhalten von vielen "niedrigeren" abhängen, d.h., letztere treten als Parameter in den Gleichungen der ersteren auf, aber nicht umgekehrt. Solche Systeme sind bottom-up gerichtet, d.h. die Dynamik auf der höheren Ebene wird von der Dynamik auf der tieferen Ebene mitbestimmt, aber nicht umgekehrt. Umgekehrt kann man eine top-down-Kontrolle des Gesamtsystems modellieren, indem man die Dynamik des Gesamtsystems durch die Dynamik einiger weniger Zustandsvariable bestimmt sein läßt. Das kann hinwiederum durch verschiedene mathematische Effekte hervorgerufen werden (etwa *centralization* bei von Bertalanffy *Bio* 1968, Versklavungsprinzip in der Synergetik, vgl. Haken *Phys* 1983). Die "dominierenden" Variable können einer "höheren" Organisationsebene zugewiesen werden; ihre Dominanz entspricht einer top-down-Organisation des Gesamtsystems.

Oft möchte man erklären, wie es in Systemen mit interagierenden "niedrigen" Teilsystemen auf einer höheren Ebene durch Selbstorganisation zu einer Ordnungsbildung kommt. Für die KI relevant sind z.B. Musterbildungsprozesse in erregbaren (z.B. neuronalen) Medien oder die Koordination/Kontrolle derFreiheitsgrade eines motorischen Systems (vgl. Einführung in Smith & Thelen 1993a). Selbstorganisation und Ordnung sind keine wohldefinierten Begriffe. Verschiedene inzwischen klassische Ansätze haben verschiedene Aspekte dieses intuitiv umschriebenen Phänomenbereichs präzisiert und analysiert: das Versklavungsprinzip von Haken (*Phys* 1983), das Prinzip "order from fluctuations" der Brüsseler Schule (Prigogine *Phys* 1980), der Hyperzyklus von Eigen und Schuster (*Bio* 1977f). Diese Ansätze betreffen kontinuierliche Systeme. Bei diskreten Systemen (z.B. Classifier-Systeme, Zellularautomaten) untersucht man, wie sich aus dem kollektiven Verhalten vieler "kleiner" Systemkomponenten durch lokale Interaktion eine globale Ordnungsbildung ergibt. Hierbei werden gerne Methoden der Informationtheorie bzw. der statistischen Mechanik verwendet. Manchmal wird für dieses sehr reiche und vielfältige Gebiet der Begriff der "emergent computation" gebraucht. Eine Sammlung bietet Forrest (1990).

Systemtheoretische Methoden sind auch hilfreich, wenn man hierarchische Systeme **über mehrere Organisiationsebenen hinweg einheitlich beschreiben** möchte. Unter einem hierarchischen System verstehe ich hierbei ein System, das in mehreren Organisationsebenen geschichtet ist.

In der logikorientierten KI ist eine einheitliche Beschreibung mehrerer Organisations-ebenen nicht üblich. Erstens werden in der Regel überhaupt nur die höheren, der rationalen Informationsverarbeitung zuzurechnenden Ebenen erfaßt. Zweitens bestehen kategoriale Unterschiede zwischen den Entitäten, die auf verschiedenen Ebenen anzutreffen sind. So sind etwa Attribute aus logischer Sicht etwas wesentlich anderes als Konzepte, und diese hinwiederum sind in ihrer Natur verschieden von noch höher angesiedelten Entitäten wie etwa Propositionen oder Inferenzregeln.

Diese kategoriale Unterscheidung birgt Ungereimtheiten. Schon Smith & Medin (*Ling*, *Kog* 1981) führen das Beispiel an, daß das perzeptuelle Feature des mittleren Querbalkens im Buchstaben E Konzeptstatus annehmen kann, wenn man die visuelle Inspektion darauf fokussiert. Medin & Barsalou (*Kog* 1987) vergleichen ausführlich empirische Befunde über *sensory perception categories* (das wären z.B. Farbeindrücke, Kantensegmente u.ä.) und *generic knowledge categories* (das sind die klassischen Konzepte wie Ball, Haus oder Kindergeburtstag). Sie stellen fest, daß beide Arten von mentalen Einheiten im großen und ganzen dieselben Eigenschaften haben. Weitere Argumente für die Ununterscheidbarkeit finden sich bei Jaeger (*KI*, *Math* 1994a, 131ff).

Es wäre denkbar, den empirischen Befund "Konzept ~ Attribut" durch die Verwendung typenfreier Repräsentationsformalismen einzufangen. So könnte man den Lambda-Kalkül bzw. kombinatorische Algebren verwenden (Einführung: Engeler *Math* 1983), und als modelltheoretischen Rahmen eine nichtfundierte Mengenlehre (Aczel *Math* 1988, Einführung in Barwise & Moss *Math* 1991). Meines Wissens wird das nirgends getan, und ich habe einen eigenen kleinen Anlauf in diese Richtung (Jaeger *KI*, *Math* 1992) nicht weiter ausgebaut.

In der Praxis wird die theoretisch geforderte kategoriale Trennung der Ebenen gern verwischt. Kommerzielle Werkzeuge zur Wissensrepräsentation, wie etwa KEE oder Knowledge Craft, erlauben einen Durchgriff auf die zugrundeliegende Programmiersprache und damit eine logische Anarchie. Auch umfassendere, empirisch motivierte, komplexe symbolische Modelle kognitiver Informationsverarbeitung wie etwa das "dynamic memory" von Schank (*Kog, KI* 1982) oder der schon mehrfach erwähnte "schema mechanism" von Drescher (*KI* 1991) enthalten eine Vielzahl interagierender Konstrukte, für die eine logische Typisierung unmöglich erscheint und im übrigen auch von niemandem versucht wird.

Die Lage in der klassischen, logikorientierten Konzeptmodellierung ist also von einem unaufgelösten Konflikt geeprägt, nämlich zwischen der theoretisch geforderten kategoriellen Verschiedenheit der Attribut- und der Konzeptebene (und anderen Ebenen) einerseits, und der empirischen Ähnlichkeit bzw. praktischen Vermengung solcher Ebenen andererseits.

Gekoppelte Differentialgleichungen oder Diffeomorphismen ermöglichen demgegenüber eine direkte, einheitliche Modellierung mehrerer Organisationsebenen und deren Interaktionsdynamik. Das Grundrezept ist einfach (vgl. Abb. 12). Man überlege sich, wieviele Ebenen (in Abb. 12 drei) und welche Observable auf diesen Ebenen man modellieren möchte (hier z, y_i, x_j), und welche Observable welche andere dynamisch beeinflussen (Pfeile). Dann muß man "nur" noch für jede der Variable eine Differentialgleichung oder einen Diffeomorphismus finden, der die beeinflussenden anderen Variable als Parameter enthält und die gewünschte Art des Einflusses ausdrückt. Ein Diffeomorphismus für y_2 zum Beispiel hätte laut den Vorgaben der Architekturskizze in Abb. 12 die Form $y_2(t + 1) = f(x(t), z_2(t), z_3(t), z_4(t))$. Durch die spezielle Form von f sind für hierarchischen Systemen typische Effekte zu erzielen, etwa daß die Dynamik der hierarchisch höheren Observable y_2 gegenüber der Dynamik der niedrigeren Variablen z_2, z_3, z_4 langsamer ist.

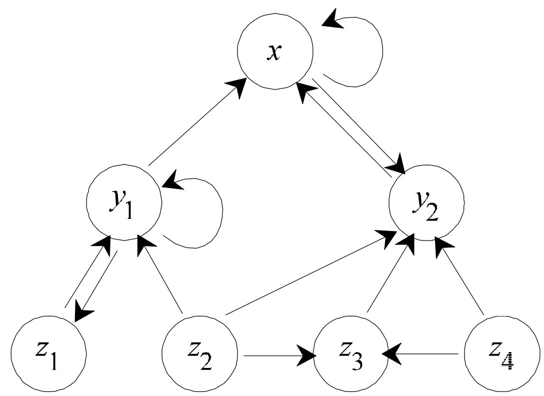

Abb. 12: Architekturskizze eines hierarchischen Systems.

In Abb. 12 kommen sowohl auf- als auch absteigende Pfeile vor, was einer simultanen Modellierung von top-down und bottom-up-Interaktionen zwischen den Ebenen entspricht. Daß Abb. 12 wie ein neuronales Netz aussieht, ist kein Zufall. Mehrschichtige neuronale Netze sind wohl der am intensivsten beforschte Typ hierarchischer Systeme. Die hierzu vorhandene Literatur ist uferlos. Es lohnt sich vielleicht, darauf hinzuweisen, daß man sich in der neuronalen-Netz-Forschung auf wenige Arten von hierarchischen Strukturen à la Abb. 12, und auf spezielle Klassen von dynamischen Gesetzen beschränkt. Die allgemeine Theorie dynamischer Systeme erlaubt demgegenüber die Modellierung von hierarchischen Systemen in einer Vielfalt, die bisher noch nicht ansatzweise ausgelotet ist.

Auch diskrete dynamische Systeme können hierarchisch aufgebaut werden.

Jedem KI-Forscher ist vertraut, daß man Algorithmen bzw. Automaten (von endlichen Automaten bis zu Turingmaschinen) im Sinne von geschachtelten Unterprogrammen hierarchisch verknüpfen kann. Im technischen Sinne (etwa der abstrakten Systemtheorie von Zadeh *Math*, *Rob* 1969 und der klassischen Kybernetik von Wiener *Math* 1961[2]) beschreiben alle geschachtelt rekursiven Computerprogramme hierarchische dynamische Systeme. Diese Beobachtung ist jedoch nicht weiter hilfreich, denn in der KI, der Linguistik und der mathematischen Algorithmentheorie bemüht man sich inzwischen hauptsächlich um eine logikorientierte, an formalen Semantiken ausgerichtete Erforschung solcher "Systeme", unter Vernachlässigung der zeitlichen Komponente. Zwar werden in geringerem Umfang Automaten und Algorithmen auch genuin als Zeitreihen-generierende dynamische Systeme erforscht (z.B. Crutchfield *Phys*, *Math* 1990, Crutchfield & Young *Phys*, *Math* 1990, Tani et al. *NN*, *Inf* 1995). Hier sind bislang jedoch meines Wissens noch keine hierarchisch geschachtelten Systeme untersucht worden.

Der Wunsch nach Schließung dieser Lücke war ein Motiv bei meiner Entwicklung der "dynamischen Symbolsysteme" (Jaeger *KI*, *Math* 1994a). Diese Systeme sind sowohl diskret und hierarchisch, als auch genuin dynamisch. Höhere Organisationsebenen gründen hier in den tieferen nicht durch eine Schachtelung im Sinne von Unterprogrammen, sondern durch eine mehr oder weniger feste Kopplung ansonsten auch autonom laufender Subsysteme, ähnlich wie in gekoppelten Differentialgleichungen oder Diffeomorphismen. Die Erforschung dieser Systeme steckt aber noch in den Kinderschuhen.

Die einheitliche Beschreibbarkeit mehrerer Organisationsebenen ist einer der Gründe, warum dynamische Systeme für Kognitionswissenschaftler und Linguisten interessant sind. Drei Beispiele sollen hier angeführt werden. Köhler (*Ling* 1987) untersucht Sprachverstehens-prozesse unter verschiedenen Gesichtspunkten der Selbstorganisation. Er interpretiert eine Vielzahl von Observablen, die aus logischer Sicht höchst unterschiedlichen Typs sind, als numerische Variable und setzt sie dann über Differentialgleichungen bzw. deren Lösungen miteinander in Bezug. Rickheit & Strohner (*Ling* 1992) stellen in einem methodologischen Aufsatz unter dem Stichwort "Integrität kognitiver Systeme" heraus, daß (unter anderem) sensorimotorische, syntaktische, semantische und pragmatische Prozesse beim Sprachverstehen inseparabel gekoppelt sind. Tucker & Hirsh-Pasek (*Ling*, *Psych* 1993) behandeln die Entwicklung des Sprachverstehens beim Kind als einen selbstorganisierenden Prozeß, in welchem höhere bzw. spätere Fähigkeiten aus früheren emergieren. Dabei werden — aus traditioneller Sicht — heterogene Typen von Beschreibungsgößen in eine Entwicklungsreihe eingeordnet. Beispielsweise wird aufgezeigt, wie einfache syntaktische Strukturen aus phonetischen Regularitäten hervorgehen.

Wenn man die phylogenetische und/oder ontogenetische Entwicklung intelligenter Informationsverarbeitung als selbstorganisierenden Prozeß ansieht, bei dem die konzeptuelle Verarbeitungsebene ohne kategorialen Typensprung aus früheren Ebenen emergiert, so kann man zu einer radikalen Konsequenz geleitet werden. Man kann dann nämlich das Sprach-vermögen als eine Art Epiphänomen sehen, das ohne besonderen entwicklungsgeschichtlichen Mehraufwand aus einem bereits funktionierenden konzeptuellen System hervorgeht. Diese Sicht wird etwa von Wildgen (*Ling* 1985, p. 96) vor dem Hintergrund einer system-theoretischen Sprachtheorie vertreten. Interessant ist in dieser Hinsicht auch Studie von Greenfield (*Psych*, *Bio* 1991). Sie analysiert onto- und phylogenetische, neurophysiologische und verhaltensbiologische Befunde und gelangt zu der Auffassung, daß beim Menschen die Fähigkeiten zu hierarchischer und sequentieller Organisation von Aktions- *und* Spracheinheiten homolog sind, d.h. in *denselben* neuronalen Strukturen gründen.

Die einheitliche Beschreibbarkeit mehrerer Organisationsebenen läßt auch ein klassisches Problem der KI in neuem Licht sehen, nämlich die Beziehungen zwischen **analogen** und **propositionalen** Repräsentationen (Sloman *KI* 1975, Hayes *KI*, *Phil* 1985). Viele Schwierigkeiten hierbei hängen mit der Logikorientiertheit der klassischen KI zusammen. Logische Repräsentationen sind ihrer Natur nach propositional. Will man auch analoge Repräsentationen integrieren, muß man entweder logische Formalismen zweckentfremden oder ein zusätzliches, analoges Repräsentationsformat in sein KI-Programm einbinden. Beides sind keine eleganten Lösungen.

Modelliert man ein intelligentes System als dynamisches System, so beschreibt man es in einem "primitiven", einheitlichen Format, in welchem gewisse Phänomen sekundär als analoge oder propositionale Repräsentationen interpretiert werden können. Dies ist allerdings eher ein Forschungsprogramm als ein bereits erreichtes Forschungsresultat. Analoge und propositionale Repräsentationen sind zwar beide schon, je für sich, im "primitiveren" Substrat dynamischer Systeme verwirklicht worden (analoge: z.B. somatotopische Repräsentation einer Hand in einem Kohonen-Netzwerk bei Obermayer et al. *NN* 1990b, Repräsentation von räumlichen Hindernissen in einem dynamischen skalaren Feld bei Engels & Schöner *AA* 1995; propositionale: vgl. die Arbeiten zur Integration symbolischer und neuronaler Prozesse im Sammelband Bookman & Sun *KI*, *NN* 1993, insbesondere Mani & Shastri *KI*, *NN* 1993). In ein dynamisches Gesamtsystem integriert worden sind beide Aspekte aber bisher wohl noch nicht. Das Repräsentationsformat im Formalismus der dynamischen Symbolsysteme (Jaeger *KI*, *Math* 1994a) sind gelabelte gerichtete Graphen. Sie erlauben eine direkte Interpretation sowohl als propositionale wie auch als analoge Repräsentationen. Der Formalismus besteht aber vorerst nur auf dem Papier.

Man kann sogar noch weitergehen und bei Tieren und Robotern neben analogen und propositionalen Repräsentationen noch eine dritte Sorte von "Repräsentationen" ins Spiel bringen, nämlich den Körper selbst. Auch dieses "Format" gilt es mit den beiden anderen zu einer Ganzheit zu integrieren, was wiederum am ehesten mit systemtheoretischen Mitteln denkbar ist (Dautenhahn & Christaller *AA*, *KI* 1995).

Die wichtigsten Punkte einer systemtheoretischen Sicht auf Organisationsebenen seien noch einmal zusammengefaßt:

- Die Emergenz einer höheren Organisationsebene über einer niedrigeren ist ein zentraler systemtheoretischer Untersuchungsgegenstand. Der schillernde Begriff der Selbstorganisation wird sehr oft in diesem Sinne verwendet. Es handelt sich hier nicht um ein einziges, klar umrissenes Phänomen, sondern um eine Vielfalt von Erscheinungen.
- In logikorientierten Modellen informationsverarbeitender Systeme werden Übergänge zwischen Organisationsebenen *beschrieben*. Systemtheoretische Modelle setzen demgegenüber nicht auf der phänomenalen Oberfläche an, sondern an den systemerhaltenden und -generierenden Mechanismen. Sie können solche Übergänge *erklären*.
- Logikorientierte Modelle verwenden typischerweise in verschiedenen Organisationsebenen verschiedene logische Typen. Demgegenüber können mit Differentialgleichungen bzw. Diffeomorphismen hierarchische Systeme in einem homogenen Format modelliert werden. Das gilt im Prinzip auch für diskrete Systeme, wo es aber noch kaum eine einschlägige Forschungstradition gibt.
- Analoge und propositionale Repräsentationen lassen sich als abgeleitete Phänomene in basaleren, allgemeineren dynamischen Systemmodellen interpretieren.
- Systemtheoretische Modelle erlauben insgesamt ein vertieftes und differenziertes Verständnis der Integrität von Systemen. Die konzeptuelle Verarbeitungsebene hat nicht den Sonderstatus wie in der logikorientierten KI.

4.2.4 Agent-Umwelt-Systeme

Körperlichkeit und physikalische Situiertheit

Die Natur vieler Intelligenzleistungen kann nur verstanden werden, wenn man die physikalische Körperlichkeit von intelligenten Agenten und die physikalische Interaktion mit der realen Umgebung berücksichtigt. Diese Einsicht ist in der traditionell einer entkörperlichten (engl. *disembodied*) Intelligenzauffassung verbundenen KI recht jung und keineswegs ausdiskutiert. In der *situated action* Debatte in der KI und der Kognitionswissenschaft (z.B. Suchman *KI* 1987, Vera & Simon *KI* 1993a, Clancey *KI* 1993) werden verschiedene Aspekte dieser Perspektive teilweise in ungewöhnlich polemischer Weise (z.B. Vera & Simon 1993b, Hayes et al. *KI*, *Phil* 1994) diskutiert. Für die behavior-orientierte Robotik ist die Körperlichkeit und physikalische Situiertheit von Robotern der zentrale methodologische Ausgangspunkt (Brooks *KI*, *AA*, *Rob* 1991, Steels *KI*, *AA* 1993).

Wenn man formale Modelle körperlicher und physikalisch situierter Agenten aufstellen will, bieten sich systemtheoretische Techniken kanonisch an.

Was den Aspekt der Körperlichkeit betrifft, so *sind* Tier- und Roboterkörper mechanische, dynamische Systeme. Dank ihrer geschichtlichen Herkunft aus der physikalischen Mechanik sind kontinuierliche dynamische Systeme von Hause aus als formale Modelle prädestiniert. Dies ist bereits in 4.2.2 ausgeführt worden.

Ich möchte hier etwas ausführlicher auf den zweiten Aspekt eingehen, den der physikalischen Situiertheit. Obwohl physikalische Situiertheit ein Leitmotiv für die Paradigmen der *situated action* und der behavior-orientierten Robotik ist, so ist noch nicht ganz klar, was dieser Terminus eigentlich bedeuten soll und welche Konsequenzen die damit verbundenen Ideen für die Erforschung der Intelligenz haben. Hier sollen einige Aspekte dieses Komplexes aus systemtheoretischer Sicht näher beleuchtet werden.

Eine besonders konsequente Version des Situiertheitsgedankens besteht darin, Agent und Umwelt als ein einziges dynamisches System aufzufassen. Dies wird besonders von Vertretern der behavior-orientierten Robotik getan (erste Skizze in Steels *KI*, *AA* 1993b, etwas formalere Skizzen in Smithers *AA* 1994, Beer *KI*, *AA*, *NN* 1995). Hier wird vorgeschlagen, Agent-Umwelt-Systeme durch Differentialgleichungen zu beschreiben, wobei der Agent und die Umwelt jeweils formal als ein Teilsystem aufgefaßt werden.

Diese Arbeiten sind Prinzipienpapiere. Zwar lassen sich aus der Idee, Agenten-Umwelt-Systeme als kontinuierliche dynamische Systeme aufzufassen, interessante Folgerungen bzw. Forschungsfragen ableiten (z.B. Bedeutung von Attraktoren, Instabilitäten, etc.). Möglicherweise lassen sich auch einfache Roboter noch als offene Systeme durch Differentialgleichungen modellieren. Was die Modellierung der Umwelt betrifft, so zeigen die von Sacks (*KI* 1990, 1991) vor dem Hintergrund des *qualitative reasoning* beschriebenen Techniken zur Darstellung und Analyse kontinuierlicher physikalischer Systeme zwar, wie man im Prinzip physikalische Umwelten mit differentialgeometrischen Techniken darstellen und analysieren kann. Seine Arbeiten betreffen aber lediglich zweidimensionale Systeme. Es scheint mir ausgeschlossen, auf diese Weise eine hochdimensionale nichttriviale Umwelt zu modellieren. Nur zwei von vielen Gründe möchte ich hier für diese Vermutung angeben. Erstens müßte man zur Berücksichtigung räumlicher Gegebenheiten partielle Differentialgleichungen verwenden, was die Sache mathematisch sehr kompliziert. Zweitens sind Umwelten meistens vorgegeben und nicht konstruiert, d.h. man hat die waltenden dynamischen Gesetzmäßigkeiten nicht als Schöpfer des Systems in der Hand und kann sie kaum kennen.

So stellt sich die Frage nach vereinfachten, diskret approximierenden, evtl. stochastischen Systemmodellen insbesondere für die Modellierung der Umwelt in einem Agent-Umwelt-System. In diese Richtung gehen Kaelbling und Mitarbeiter (Entwurf und Motivation in Kaelbling *KI* 1991, Ausführung in Baayse et al. *KI* 1995). Sie fassen die Umwelt als endlichen Automaten auf, in dem Aktionen eines Roboters für Zustandsübergänge sorgen. Für einen Roboter stellt sich die Aufgabe, unter Bedingungen sensorischer und aktuatorischer Unsicherheit aus Aktionssequenzen ein Modell der Umwelt zu lernen, d.h. ein endliches Automatenmodell zu generieren. Die Autoren geben hierfür praktikable Algorithmen an. Es wäre eine spannende Aufgabe, solche diskreten Umweltmodelle mit formal gleichartigen Agentenmodellen (z.B. Košecká & Bajcsy *Rob* 1994) zu umfassenden diskreten Modellen von Agenten-Umwelt-Systemen zu integrieren.

Wenn auch die systemtheoretische Modellierung von kompletten Agent-Umwelt-Systemen gegenwärtig noch nicht erreicht ist, so kann man immerhin gedanklich den Agenten als Teilsystem eines solchen umfassenden Systems erfassen und zunächst für sich modellieren. Die Interaktionen zwischen Agent und Umwelt müssen dann gleichsam "abgeschnitten" werden, und die "Schnittstelle" muß geeignet modelliert werden. Das führt dazu, Agenten als *offene Systeme* aufzufassen.

Das ist ein schillernder Begriff, mit dem seit jeher schwierige epistemologische und mathematische Fragen verbunden sind. Im engeren physikalischen Sinn wird ein System als offen bezeichnet, wenn es Materie und/oder Energie mit der Umgebung austauscht. Unter dem Stichwort "Selbstorganisation fern vom Gleichgewicht" hat hier besonders die Brüsseler Schule (Prigogine *Phys*, *Math* 1980) exakte, neuartige Einsichten erschlossen. In unserem Zusammenhang ist hervorzuheben, daß Lebewesen (vgl. von Bertalanffy *Bio* 1968 Kap. 6) und Ökosysteme (vgl. Zwölfer *Bio* 1986) in diesem strengen Sinn offene Systeme sind und auch mit entsprechenden exakten Methoden modelliert werden. Der Energiehaushaltes eines

Agenten bzw. die Gesamt-Energiebilanz eines Agent-Umgebungssystems werden auch in den von McFarland angestoßenen Forschungen im Robotik-Labor von Luc Steels als Modellierungsfaktoren betont.

Der Begriff des offenen Systems wird darüberhinaus auch mehr oder weniger nach Belieben für andere Austauschprozesse eines Systems mit seiner Umgebung verwendet, etwa informationeller oder monetärer. Solche Varianten sind dann in der Regel weniger formalisiert, aber für das Verständnis psychischer, sozialer oder wirtschaftlicher Systeme nützlich (Reader: Kornwachs *Phil*, *Phys* 1984). Die ebenfalls wenig formale und nicht auf materielle/energetische Austauschprozesse fixierte Theorie autopoietischer Systeme betont übrigens gerade ein der Offenheit entgegengesetztes Prinzip, nämlich die "operationalen Geschlossenheit" lebender und sozialer Systeme (vgl. Willke *Soz* 1987, Kap. 3.2, Maturana & Varela *Bio*, *Phil* 1984), was hinwiederum zu Schwierigkeiten bei der Erklärung der Anpassung eines Agenten an seine Umwelt führt (Roth *Bio*, *Phil* 1986).

Für die KI und die Robotik reduziert sich das Problem der Offenheit eines Agenten rasch auf eine geeignete Modellierung seiner Wahrnehmungsfunktionen. Mathematisch scheint auf den ersten Blick klar zu sein, wie das im Prinzip zu geschehen hat. Bei kontinuierlichen Systemmodellen, die durch Differentialgleichungen $d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}, \mathbf{u})$ bzw. Diffeormorphismen $\mathbf{x}(t + 1) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t))$ spezifiziert werden, dienen freie zeitabhängige Parameter $\mathbf{u}$ zur Modellierung von sensorischem Input. Auf diese Weise ist insbesondere sensorischer Input in neuronale Netzwerke, der verbreitetsten Klasse solcher Systeme, formal verwirklicht.

Hier ergibt sich aber ein weitgehend ungelöstes Problem. Technisch gesehen sind die Parameter $\mathbf{u}$ Kontrollparameter. Das qualitative Verhalten des Systems $d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}, \mathbf{u})$ bzw. $\mathbf{x}(t + 1) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t))$ kann aber mit den üblichen Techniken nur analysiert werden, wenn die Kontrollparameter als Konstanten angenommen werden oder sich zum mindestens im Vergleich zur Systemdynamik nur sehrlangsam verändern. Tatsächlich werden neuronale Netze mit Input meist unter dieser Annahme betrieben. Typischerweise werden Inputwerte festgehalten und dem Netz Zeit gelassen, einen stabilen Antwortzustand einzunehmen (der durchaus eine chaotische Aktivität sein kann, etwa bei Yao & Freeman *NN*, *Bio* 1990).

Diese Annahme entspringt aber nur unserem mathematischen Unvermögen, mit *schnell* veränderlichen freien Parametern $\mathbf{u}$ umzugehen. Inhaltlich gerechtfertigt ist sie nicht. Ganz im Gegenteil, die auf einen Agenten wirkende Sensorinformation wird im allgemeinen in derselben zeitlichen Größenordnung veränderlich sein wie manche interne Zustandsgrößen selbst. Die Schwankungen in den Lesewerten eines Infrarotsensors bei einem mobilen Roboter etwa verändern sich gerade durch, und im gleichen Zeitmuster mit, den Bewegungen des Roboters selbst, die hinwiederum direkt mit internen Zustandsgrößen gekoppelt sind. Formal führt das auf das Problem, Systeme der Art $d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}, \mathbf{u})$ bzw. $\mathbf{x}(t + 1) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t))$ zu analysieren bzw. zu konstruieren, wo die Dynamik $\mathbf{u}(t)$ unbekannt und schnell ist. Da die eigentliche Systemdynamik $\mathbf{x}(t)$ durch die externe Dynamik von $\mathbf{u}(t)$ beliebig deformiert werden kann, ist hier ohne geeignete einschränkende Annahmen keine Theoriebildung möglich.

Es gibt hier noch nicht viel zu berichten. Unter Ergodizitätsannahmen über $\mathbf{u}(t)$ definiert Casdagli (*Phys*, *Math* 1992), was chaotisches Verhalten für solche stochastisch getriebenen, offenen Systeme bedeutet. In Jaeger (*Math*, *Rob* 1995) schränke ich den Variationsbereich von $\mathbf{u}$ und die Klasse von $\mathbf{f}$ ein, um den Erhalt des strukturellen Typs von $\mathbf{f}$ bei schneller Variation von $\mathbf{u}$ zu gewährleisten. Dies ist interessant für ein modulares Design von Kontrollprogrammen für behavior-basierte Roboter. Einzelne Behaviors können als offene (gegenüber anderen Behaviors und gegenüber Sensordaten) Systeme entworfen werden; werden sie gekoppelt bzw. mit Sensordaten gefüttert, so können sie in ihrem Verhalten zwar moduliert werden, jedoch bleibt der qualitative Typ der Moduldynamik erhalten.

In der *situated action*-Debatte ist die Frage nach der Funktion, oder gar der Existenz von internen Repräsentationen aufgekommen. In der klassischen KI und der Kognitionswissenschaft spielen bekanntlich symbolische Repräsentationen und syntaktische Operationen darauf eine zentrale Rolle. Dies entspricht einer rationalen Informationsverarbeitung. Demgegenüber wird von manchen Vertretern der *situated action* die direkte, sensomotorische Rückkopplung vom Agenten in seine Umgebung in einer Weise interpretiert, welche die Existenz von symbolischen Repräsentationen überhaupt infrage stellt. Diese Debatte über die Natur von Repräsentationen in situierten Agenten ist heftig, verwirrend und unabgeschlossen (Einstieg etwa durch das Sonderheft 17 No 1 (1993) von Cognitive Science oder die Proceedings der Konferenz "On the Role of Dynamics and Representation in Adaptive Behavior and Cognition", vgl. Referenz zu Smithers 1994). Es bestehen diffuse Berührungen mit dem Thema dynamische Systeme:

- Kritiker des klassischen Repräsentationalismus befürworten oft, Agenten als dynamische Systeme aufzufassen (z.B. Clancey *KI* 1993, Pfeifer *Rob* 1993).
- In der *symbol grounding*-Debatte, die mit der *situated action*-Debatte zusammenhängt, und in der Debatte um verteilte Repräsentationen in neuronalen Netzwerken, die hinwiederum mit der *symbol grounding*-Debatte zusammenhängt, wird die kontinuierlich-dynamische Natur symbolischer Repräsentationen hervorgehoben (z.B. Dalenoort *Psych* 1990a,b, Scheerer *Psych* 1990, Chalmers *Phil* 1990, 1992, MacLennan *Inf* 1991, van Gelder & Port *Kog* 1994a).
- Die Position der *situated action* ist gut verträglich mit einer interaktionistischen Auffassung von Konzeptungen, welche hinwiederum (s.o.) zu einer systemtheoretischen Sichtweise paßt.

Ich möchte hier zu dieser philosophischen Debatte nichts weiter ausführen. Es kann aber nicht schaden, ausdrücklich festzustellen, daß eine systemtheoretisch inspirierte Sicht der Dinge nicht nur gut mit der Existenz von internen Repräsentationen verträglich ist, sondern sogar zu einer besonders reichen und interessanten, nämlich dynamischen Auffassung von Repräsentationen führen kann — was hoffentlich die vorliegende Arbeit reichlich illustriert.

Multi-Agenten-Systeme

Oft enthält die Umwelt eines Agenten weitere Agenten, oder sie besteht sogar in erster Linie aus solchen. Für die KI liegt dieser wichtige "Spezialfall" in mindestens drei Varianten vor: bei Mensch-Maschine-Kommunikation, bei Roboter-Roboter-Interaktionen, und in der agentenorientierten Programmierung/verteilten KI (Einführung: Müller *KI* 1993). Wo können hier systemtheoretische Methoden helfen?

Die Betrachtung von Multi-Agenten-Systemen im weiteren Sinn findet man schon bei den Klassikern der Systemtheorie (vorsichtige Bemerkungen über Information und Gesellschaft bei Wiener *Math*, *Bio* 1961, allgemeine systemtheoretische Betrachtungen über Humanwissenschaften und Psychiatrie bei von Bertalanffy *Bio* 1968). Der Begriff der autopoietischen Systeme hat seine größte Bedeutung vielleicht erst in den Sozialwissenschaften erlangt (Willke *Soz* 1987). Naturwissenschaftlich-synergetische systemtheoretische Betrachtungsweisen haben in die Wirtschaftswissenschaft (Forrester 1972) und die Managementtheorie (Ulrich & Probst 1984) Eingang gefunden. Schließlich sei noch angemerkt, daß die Populationsbiologie (Wilson & Bossert *Bio* 1973), eine der klassischen Domänen für strenge systemtheoretische Techniken, ebenfalls in gewissem (statistischen) Sinn von Multi-Agenten-Systemen handelt.

Alle diese Forschungen belegen, daß systemtheoretische Methoden prinzipiell für Multi-Agenten-Systeme anwendbar erscheinen. Für die Belange der KI bieten sie jedoch keine direkt nutzbaren Beiträge, in erster Linie weil es sich meistens um wenig formale Betrachtungen handelt.

Ich kenne nur zwei Fälle einer formalen und damit möglicherweise technisch verwendbaren, für die KI direkt interessanten Analyse von Multi-Agenten-Systemen. Der erste

Fall sind Newtsons (*Psych* 1993) Zeitserienanalysen der Kommunikationsdynamik in Mensch-Mensch-Dyaden. Die hier entwickelten Analysemethoden scheinen auf Roboter-Roboter- und Mensch-Maschine-Kommunikation übertragbar zu sein und könnten zur Evaluierung und evtl. zur Steuerung von Kommunikation zwischen wenigen Agenten genutzt werden. Der zweite Fall sind die Arbeiten von Deneubourg und Mitarbeitern, die seit langem die emergente Organisation von Insektengesellschaften mit systemtheoretischen Techniken analysieren (z.B. Deneubourg *Bio* 1975, Deneubourg et al. *Bio* 1991, Deneubourg & Franks *Bio* 1995). Diese schönen Arbeiten sind offenbarer hochrelevant für die verteilte KI, werden aber in Deutschland noch kaum rezipiert.

Vergleich mit der klassischen Sicht

Auch in der klassischen, logikorientierten KI werden Agent-Umwelt-Systeme und Multi-Agenten-Systeme beschrieben. Agent-Umwelt- und Agent-Agent-Interaktionen werden durch verschiedene nichtmonotone Logiken, vor allem Modallogiken, modelliert. Mitunter findet sich hier auch der Begriff des dynamischen Systems (etwa in der großen Studie von Sandewall *KI, Math* 1994, der Agent-Umwelt-Systeme als "inhabited dynamical systems" bezeichnet). Insgesamt ist das hiermit umrissene Gebiet des "reasoning about actions and plans" (RAAC) eine große, integrative und sehr produktive Strömung in der KI. Sie ist aber rein logikorientiert und fällt nicht mehr in den Bereich dieser Übersichtsarbeit. Die Perspektive der *situated action* und die behavior-orientierte Robotik haben sich teilweise in kritischer Auseinandersetzung mit dem RAAC-Paradigma entwickelt (vgl. z.B. Aufsätze von Chapman/Agre und Kaelbling in Georgeff & Lansky *KI* 1986). Einige wichtige Unterschiede zwischen der logikorientierten und der systemtheoretischen Sicht auf Agent-(Agent-)Umwelt-Systeme sind die folgenden:

- In der Tradition des RAAC wird *über* Systeme räsoniert, die als bestehend vorausgesetzt werden. Systemtheoretische Techniken dienen demgegenüber dazu, die Systeme *selbst* zu spezifizieren. Verkürzt gesagt, logische Beschreibungen werden von vorgegebenen Systemen passiv angetrieben, während systemtheoretische Spezifikationen die Systeme aktiv antreiben.
- In der Sicht des RAAC gibt es einen *recognize-plan-act-cycle*; insbesondere sind Beobachtung und Aktion verschiedene Ereignissorten. In systemtheoretischer Sicht bilden Wahrnehmung und Aktion dagegen eine dynamische Einheit.
- RAAC betrifft grundsätzlich eine rationale Ebene der Informationsverarbeitung. Systemtheoretische Ansätze beschäftigen sich in der Praxis zumeist mit "niedrigeren" Verarbeitungsebenen.

Die genauere Analyse dieser Unterschiede und die Suche nach möglichen Integrationen wäre meiner Meinung nach überaus lohnenswert, kann hier aber nicht versucht werden.

4.2.5 *Physics of Computation*

Ab und zu tauchen in der (physikalischen) Literatur Arbeiten auf, welche sich mit den Beziehungen zwischen *computation* und der Physik beschäftigen. Dabei geht es einerseits um die Frage, ob und wie sich die Dynamik der physikalischen Realität als *computation* verstehen läßt, und andererseits um die Frage nach physikalischen Randbedingungen von rechnenden Systemen. Da hier einerseits Methoden der statistischen Mechanik verwendet werden, die im weiteren Sinne der Systemtheorie zuzurechnen sind, und andererseits *computation* jedem KI-System zugrundeliegt, möchte ich hier der Vollständigkeit zumindest auf die Existenz solcher Arbeiten hinweisen, die freilich in keinem durchgängigen Zusammenhang miteinander stehen (z.B. Toffoli *Phys* 1991, Bennet *Phys* 1981 zur Thermodynamik der *computation*, Wolfram

Math 1983 zur statistischen Mechanik von Zellularautomaten, Wolfram *Math* 1984 zu Fragen der Unentscheidbarkeit und Intraktabilität von Problemen der theoretischen Physik, Moore *Math* 1993 zur Übertragung der Rekursionstheorie in eine kontinuierliche Dynamik).

4.3 Dynamische Symbole

In der klassischen KI, der klassischen Linguistik und großen Teilen der Kognitionswissenschaft werden informationsverarbeitende Systeme als symbolverarbeitende Systeme aufgefaßt. Was ein Symbol genau ist, ist nicht in jeder Hinsicht geklärt (vgl. Vera & Simon *KI* 1993, van Gelder & Port *Kog* 1994a, Jaeger *KI*, *Math* 1994a). Die schwierigsten Fragen hierbei betreffen die Semantik; sie sollen im folgenden ausgeblendet bleiben. Unstrittig ist aber, daß es sich bei Symbolen um physikalische Entitäten handelt, die (1) beobachtbar und effektiv identifizierbar sind (wobei probabilistische Identifizierungen nicht ausgeschlossen sind), und die (2) zu syntaktischen Strukturen komponiert werden können.

Es gibt gute Gründe für das Paradigma der Symbolverarbeitung, zumindest in seiner schwachen, technischen, von der Metaphysik der Semantik befreiten Version (1) + (2). Diese Gründe sind meiner Meinung nach so gut, daß ein systemtheoretisches Modell intelligenter Informationsvearbeitung, welches die Punkte (1) und (2) leugnet, scheitern muß. Es geht vielmehr darum, (1') in dynamischen Systemen beobachtbare, identifizierbare dynamische Phänomene zu bestimmen, und (2') dynamische Mechanismen zu finden, welche als Komposition dieser Entitäten aufgefaßt werden können.

Der einfachste Fall von (2') ist die temporale Verkettung. Es geht also im einfachsten Fall darum, in dynamischen Systemen Zeitfolgen identifizierbarer Phänomene zu beschreiben. Dies ist die minimale Grundaufgabe, die gelöst werden muß, wenn man dynamische Systeme für die Tradition der KI und der Kognitionswissenschaften fruchtbar machen will.

In diskreten Systemen (z.B. Automaten, Zellularautomaten, Classifier-Systemen) bereitet diese Grundaufgabe keine Schwierigkeit, denn dort liegen bereits symbolische Informationseinheiten vor (Zustands-, Übergangs-, Zell- oder Classifieralphabete), und die jeweilige Dynamik erzeugt auf verschiedene Weise Zeitsequenzen. Bei "feinkörnigen", kollektiven parallelen Systemen wie Zellularräumen oder Classifier-Systemen interessiert man sich allerdings meist weniger für die basale Symboldynamik auf der feinkörnigen Ebene, sondern mehr für grobkörnigere Musterbildungen, so daß auch hier die Punkte (1') und (2') auf der kollektiven Ebene anzugehen sind. Die hierbei auftretenden Probleme sind aber in vielem identisch mit den Fragen, die bei kontinuierlichen Systemen zu beantworten sind. Deswegen soll hier nur die Frage untersucht werden: *Welche Phänomene in kontinuierlichen dynamischen Systemen lassen sich als Zeitfolgen identifizierbarer Einzelphänomene interpretieren?*

Bei einer näheren Inspektion der systemtheoretischen Literatur zeigt sich, daß es sehr viele solcher Phänomene gibt! Wir müssen nur einiges aus vorangegangenen Abschnitten rekapitulieren, um uns den Reichtum an solchen Erscheinungen vor Augen zu führen.

Ein verbreiteter Gedanke ist es, Attraktoren als identifizierbare Entitäten in dynamischen Systemen anzusehen. So werden, wie wir gesehen haben, etwa Konzepte oder Behaviors aus systemtheoretischer Sicht oft durch Attraktoren modelliert. Das ist intuitiv ein anziehender Gedanke, denn Attraktoren sind per definitionem stabil, d.h. zeitlich persistent und in gewissen Grenzen unempfindlich gegen Verformung. Das erwarten wir aber gerade auch von physikalischen Symbolen.

Die scheinbar so schlüssige Gleichung "Symbol = Attraktor" erschwert jedoch leider die Antwort auf die Frage (2'). Denn eine Zeitsequenz von Attraktoren würde voraussetzen, daß Attraktoren beginnen und terminieren können. Das Beginnen ist nicht weiter schwierig — man kann leicht zu handlichen Präzisierungen davon kommen, wann eine Systemtrajektorie nach einem transienten Start von einem Attraktor "gefangen" wird. Doch wie soll ein Attraktor terminieren? Schließlich besteht die definierende Eigenschaft eines Attraktors gerade darin, die Systemtrajektorie asymptotisch, d.h. zeitlich ins Unendliche reichend, zu fixieren (Diskussion des Problems: Jaeger 1994c, Aslin 1993). Die Gleichung "Symbol = Attraktor" würde also strenggenommen bedeuten, daß in einem dynamischen System nur ein einziges Symbol, und dies für alle Ewigkeit, aktualisiert werden kann.

Es gibt glücklicherweise Mechanismen in dynamischen Systemen, die aus dieser vertrackten Lage hinaushelfen können.

Viele Lösungen für dieses Problem sind implizit in den verschiedenen in der Literatur vorgeschlagenen Mechanismen zur neuronalen Speicherung von Zeitserien enthalten (vgl. 4.2.1). Dort werden auf verschiedene Weisen stochastische Fluktuationen benutzt, um die Systemtrajektorie zum Springen zwischen verschiedenen Attraktoren zu veranlassen. Fast immer untersucht man hierbei chaotische Attraktoren, die besonders sensibel auf Störungen reagieren können.

Zufällige Schwankungen als Auslöser für ein Springen zwischen verschiedenen Attraktoren stellen einen möglichen Mechanismus für "freie Assoziationen" dar. Oft vollzieht sich jedoch der Übergang von einer identifizierbaren Einheit (Symbol, Konzept, Behavior etc.) zum nächsten nicht zufällig, sondern unter wohldefinierten Bedingungen, etwa ausgelöst durch sensorischen Input (z.B. Hinwendungsreaktion auf konditionierten Reiz) oder durch bestimmte Interaktionen eines Teilsystems mit anderen (z.B. der Einfluß des taktschlagenden Fußes auf das Summen einer Melodie). Solche nichtzufälligen Terminierungen eines Attraktors lassen sich beispielsweise durch Bifurkationen modellieren: der Attraktor wird nicht von der Trajektorie "verlassen", sondern er bricht als solcher zusammen. Formals setzt das eine Variation von Kontrollparametern voraus. Sie bringt das den Attraktor zeitweilig beherbergende (Teil-)System zu einer Bifurkation. In diesem Sinne ist wohl die katastophentheoretische Modellierung von Satzstrukturen bei Wildgen (*Ling* 1985) zu verstehen, die allerdings nicht im formalen Detail ausgeführt wird. Eine grundsätzliche Schwierigkeit bei dieser Herangehensweise ist, daß die Dynamik der Kontrollparameter auf einer ähnlichen Zeitskala anzusiedeln ist wie die Dynamik des kontrollierten Systems selbst. Diese Situation ist mit den vorhandenen mathematischen Mitteln nicht gut zu beherrschen (vgl. Abschnitt 4.2.4).

Neben Attraktoren sind auch einige andere Phänomene in dynamischen Systemen als formales Korrelat für symbolische Entitäten vorgeschlagen worden, oft von theoretischen Physikern, die Generatoren von Symbolsequenzen als dynamische Systeme beschreiben. Die einfachste dieser Modellierungen besteht darin, Symbole mit beschränkten Intervallen oder Volumina im Phasenraum zu identifizieren. Die "Produktion" eines Symbols besteht dann im Durchgang der Trajektorie durch ein solches Volumen. Dieser Ansatz ist verwandt zu der Diskretisierung des Kontinuums, wie sie in der *qualitative reasoning*-Forschung in der KI verbreitet ist (z.B. Ling & Buchal *KI* 1993). Er wird verwendet z.B. einer informationstheoretischen Analyse von Texten und Musikstücken von Ebeling & G. Nicolis (*Psych*, *Math* 1992), und einer informationstheoretischen Ableitung empirischer Worthäufigkeiten in schriftlichen Texten aus der Annahme eines zugrundeliegenden kontinuierlichen dynamischen Prozesses bei J.S. Nicolis & Tsuda (*Phys* 1988). In der letztgenannten Arbeit wird auch kurz ein weiterer Mechanismus zur Symbolsequenzproduktion in kontinuierlichen dynamischen Systemen vorgestellt. Symbole entsprechen dabei den Dimensionen des Phasenraumes, und eine Symbolproduktion dem Überschreiten einer Schwelle in der zugeordneten Dimension.

Schließlich möchte ich hier noch auf die sehr schöne Arbeit von Tino et al. (1995) hinweisen. Die Autoren trainieren rekurrente neuronale Netze auf das Simulieren von endlichen Automaten. Die Zustände des Automaten finden sich hierbei als Volumina im Phasenraum des Netzes wieder; zusätzlich werden jedoch innerhalb dieser Volumina noch Attraktoren beschrieben, welche Zustandszykeln des simulierten Automaten entsprechen. Wie in den von Tani (vgl. Abschnitt über Informationspeicherung in 4.2.1) beschriebenen Netzwerken finden auch hier während des Lernens Bifurkationskaskaden statt.

Im Gegensatz zur Verwendung von Attraktoren als symbolische Einheiten hat man bei der Verwendung von Volumina im Phasenraum nicht die Schwierigkeit, zeitliche Übergänge zwischen Symbolen zu erklären. Sie ergeben sich direkt aus der zeitlichen Entwicklung der Systemtrajektorie, die von einem Volumenelement zum nächsten läuft. Dafür ist der Aspekt der Stabilität des Symbols verlorengegangen.

Diese Sammlung von Mechanismen zur Modellierung symbolischer Prozesse in dynamischen Systemen ist unvollständig, aber doch schon recht reichhaltig. Ich finde, daß hier eine besondere Stärke systemtheoretischer Modellierung von Informationsverarbeitung deutlich wird, nämlich die Möglichkeit zu einer sehr differenzierten und präzisen Erfassung zeitlicher Phänomene. Aus der Sicht der klassischen Symbolverarbeitung werden Symbole einfach konkateniert, und mehr ist zu diesem Vorgang selbst nicht zu sagen. Demgegenüber führen systemtheoretische Modelle zur Differenzierung vieler verschiedener Symbolbildungs- und Konkatenierungsmechanismen. Sie unterscheiden sich in Aspekten, die für die Leistung und Adaptation physikalisch situierter Agenten von Belang sind, wie z.B. Geschwindigkeit, Störanfälligkeit, Interaktionsfähigkeit mit anderen Prozessen, und vielem anderen. Die Würdigung der Tatsache, daß Symbole in physikalischen Systemen notwendig *dynamische* Symbole sind, deren dynamische Aspekte durchaus von Symbol zu Symbol sehr verschieden sein können, kann meiner Meinung nach zu bedeutenden Fortschritten in der KI und den Kognitionswissenschaften führen.

4.4 Wie erklären systemtheoretische Modelle?

Naturwissenschaftliche Erklärungen sind typischerweise reduktionistisch: ein Phänomen wird durch Angabe eines "darunterliegenden" Mechanismus erklärt. Im Gegensatz dazu findet man oft die Behauptung, daß man mit systemtheoretischen Methoden zu nicht-reduktionistischen Erklärungen komme: ein Phänomen werde im Zusammenhang eines einbettenden Systems erklärt, dieses System sei aber hinwiederum durch das Phänomen mitbestimmt; d.h. man kann das Ganze und das Einzelne nur gleichzweitig und in wechselseitiger Bedingtheit verstehen. Diese Perspektive ist nicht neu; sie findet sich schon in der philosophischen Hermeneutik und der Gestaltphilosophie/Gestaltpsychologie.

Meiner Meinung nach steckt hinter solchen Aussagen erstens eine gewisse Sehnsucht nach mystischer Geborgenheit, und zweitens ein technisches Mißverständnis.

Bei näherem Hinsehen findet man, daß systemtheoretische Modelle auf zwei Arten zur Erklärung von Phänomenen verwendet werden können. Die eine Art ist schlicht und einfach die gewohnte reduktionistische. Die andere Art ist zwar in gewissem Sinne nicht-reduktionistisch, sie beruht aber auf mathematisch klaren Universalitätseigenschaften dynamischer Systeme, denen nichts Geheimnisvolles anhaftet. Außerdem sind nicht-reduktionistische Erklärungen dieser Art nur "oberflächlich" und einer reduktionistischen Fundierung fähig.

Um genauer darzulegen, wie ich mir diese beiden Erklärungsarten vorstelle, verwende ich das Beispiel eines rekurrenten neuronalen Netzes, das in "freier Assoziation", evtl. unter Einfluß von Inputparametern, Zeichenfolgen generiert. Diese werden an diskret codierenden Outputneuronen abgelesen. Dies ist ein typisches, so oder ähnlich oft in der Praxis auftauchendes dynamisches System, das dank seiner Rekurrenz, Gegliedertheit und der Beeinflussung durch externe Parameter ein reiches Repertoire an dynamischen Phänomenen, sinnvoll zu isolierenden Teilsystemen, und Bifurkationen aufweist. Wie kann man dieses empirische System mit systemtheoretischen Mitteln erklären?

Wir betrachten dazu die vier Felder in Abb. 13.

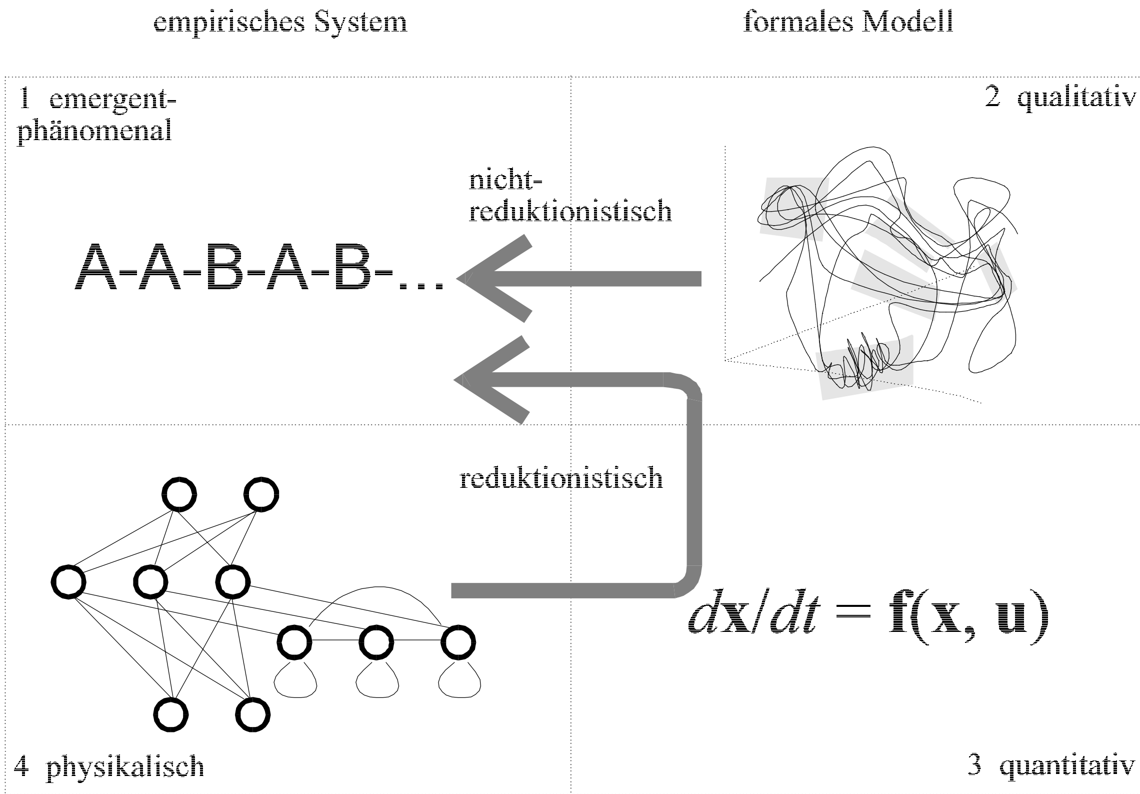

Abb. 13: Reduktionistische und nicht-reduktionistische systemtheoretische Erklärungen

Im Feld 4 ist das zu untersuchende, physikalische Systeme dargestellt. Feld 3 zeigt ein quantitatives systemtheoretisches Modell, das üblicherweise aus Differentialgleichungen (wie in Abb. 13) oder Diffeomorphismen bestehen wird. Zu ihrer Aufstellung ist eine genaue Kenntnis der systembestimmenden Variablen und deren dynamischer Zusammenhänge notwendig. In Feld 2 ist schematisch das Ergebnis einer qualitativen Analyse der Systemgleichungen dargestellt. Eine solche qualitative Analyse beschreibt den Phasenraum durch Attraktoren (möglicherweise in Teilsystemen), Bifurkationen, Übergänge der Trajektorie zwischen verschiedenen Attraktoren, und dergleichen. Im Feld 1 schließlich findet man die emergenten Phänomene, die im physikalischen System zu beobachten sind und die typischerweise in einem qualitativen Vokabular beschrieben werden.

Dieser Weg durch die Felder 4-3-2-1 entspricht einer reduktionistischen Erklärung. Das auf der qualitativen, emergenten Ebene beobachtete Verhalten des physikalischen Systems wird auf dem Umweg über ein formales Modell auf die niedrigere Ebene des physikalischen Substrats reduziert. Die Übergänge zwischen den Feldern sind alle nichttrivial. Der Übergang 4-3 beinhaltet die Aufgabe, geeignete Systemvariable zu identifizieren und deren dynamisches Verhalten zu erforschen. Diese Aufgabe ist sehr einfach, wenn man das physikalische System selbst konstruiert hat, wie das bei künstlichen neuronalen Netzen der Fall ist. Sie kann aber schwer sein, wenn man es mit einem vorgefundenen natürlichen System zu tun hat (instruktives Beispiel einer solchen Suche nach den richtigen Variablen bei Clark et al. 1993). Der Übergang 3-2 ist die mathematische Analyse der Systemgleichungen. Wie in Abschnitt 5.1 noch hervorgehoben wird, kann diese Aufgabe beliebig schwierig sein, selbst wenn man sich auf numerische Simulationen zurückzieht. Der Übergang 2-1 schließlich bedeutet eine Abbildung von formalen systemtheoretischen Phänomenen auf reale. Beispielsweise kann es (wie im Abschnitt 4.3) darum gehen, Attaktoren oder Volumina im Phasenraum auf Konzepte oder Behaviors abzubilden.

Musterbeispiele solcher reduktionistischen Erklärungen findet man bei Yao & Freeman (1990) und Ewert & v. Seelen (1974).

Bei nicht-reduktionistischen Erklärungen erklärt man emergente Phänomene im realen System durch qualitative Eigenschaften von formalen Systemen. Man beschränkt sich also auf den Übergang 2-1. Diese Art der systemtheoretischen Erklärung ist außerhalb der Naturwissenschaften die verbreitetste. Sie kann mehr oder weniger präzise durchgeführt werden. Ein Beispiel für eine präzise Durchführung ist die Deutung der spontanen Bewegungen eines Neugeborenen als chaotischer Attraktor bei Robertson et al. (1993). Hier wird die fraktale Dimension des Attraktors aus empirischen Daten berechnet. Das ist möglich, ohne ein formales Modell im Sinne des Feldes 2 (Abb. 13) zu besitzen, d.h. insbesondere, ohne sich darum zu kümmern, welches eigentlich die systembestimmenden Variablen sind. Aus der bloßen Tatsache, daß ein chaotischer Attraktor mit gewissen mathematischen Eigenschaften vorliegt, lassen sich nichttriviale Schlußfolgerungen (z.B. Stabilität, Dimensionalität eines Einbettungsraumes) ziehen. Ein Beispiel für eine weniger präzise, aber gleichwohl ansprechende Durchführung dieses Erklärungstyps findet sich bei Tucker & Hirsh-Pasek (1993). Sie nehmen an, daß das Sprachsystem beim Kind als dynamisches System zu verstehen ist, stützen diese Annahme durch eine Reihe von Beobachtungen (ohne dabei ins mathematische Detail zu gehen), und können dann z.B. Diskontinuitäten in der Sprachentwicklung als Bifurkationen erklären.

Der Erklärungsgehalt solcher nicht-reduktionistischer Erklärungen beruht in den *universellen* Eigenschaften dynamischer Systeme. Wenn die Mitglieder in einer bestimmten Klasse von dynamischen Systemen eine bestimmte qualitative Eigenschaft haben, dann braucht man zur Erklärung eines emergenten Phänomens (in Feld 1) die genaue Form des dynamischen Modells (Feld 3) nicht zu kennen, um diese qualitative Eigenschaft als Erklärung des Phänomens zu verwenden. Man muß nur belegen, daß das fragliche empirische System der betreffenden Klasse von dynamischen Systemen angehört, und kann dann dem empirischen System die betreffende qualitative Eigenschaft zuschreiben. Diese Art der Erklärung funktioniert deswegen so gut, weil die mathematische Forschung sehr interessante qualitative Eigenschaften in sehr großen Klassen dynamischer Systeme gefunden hat (z.B. Attraktortypen, Bifurkationstypen, Systemverhalten in der Nähe von Bifurkationen, slaving, gekoppelte Oszillationen, Chaos).

Nicht-reduktionistische Erklärungen dieser Art sind nicht-tautologisch, da man aus der Erkenntnis, daß eine bestimmte qualitative Eigenschaft aus einer bestimmten Klasse dynamischer Systeme vorliegt, Voraussagen ableiten kann, die man dann wieder empirisch prüfen kann, was eine Falsifikation der ursprünglich angenommenen Erklärung möglich macht. Beispiele hierfür finden sich bei Wolff (1993).

Auch wenn solche nicht-reduktionistischen Erklärungen mithin echte Erklärungen sind, darf man nicht vergessen, daß sie im Prinzip um ein formales Modell (Feld 3) und eine Rückführung auf das physikalische Substrat des empirischen Systems (Feld 4) erweitert werden können, wodurch sie zu reduktionistischen Erklärungen werden und ihre Auflösung stark zunimmt. Nicht-reduktionistische Erklärungen sind Provisorien!

Ich vermute, daß in der populären bzw. außerwissenschaftlichen Diskussion um systemtheoretische Modelle häufig "nicht-reduktionistische Erklärung" mit "Rückkopplung" verwechselt wird. Die verschiedenen Teilsysteme und Systemvariable eines dynamischen Systems sind in der Regel auf zyklischen Interaktionspfaden miteinander rückgekoppelt. Daher kann man die Dynamik eines Teilsystems oder einer Variable nicht verstehen, ohne gleichzeitig die Dynamik der anderen mitzumodellieren. In diesem Sinne sind in rückgekoppelten Systemen die Teile nur mit dem Ganzen gemeinsam zu erklären. Diese Einsicht drückt jedoch nur eine einschränkende Randbedingung systemtheoretischer Erklärungen aus; sie stellt nicht schon ein Erklärungsprinzip selbst dar.

5 Kritische Bewertung

Im vorangegangenen Kapitel wurden die Beiträge systemtheoretischer Modelle zum Verständnis intelligenter Informationsverarbeitung positiv gewürdigt. Sie seien hier noch einmal in Kürze aufgelistet:

- Ein systemtheoretisches Modell eines Agenten kann gleichzeitig sowohl eine quantitative Beschreibungen des körperlich-mechanischen oder neuronalen Apparats beinhalten wie auch qualitative, emergente Phänomene erfassen, die zur Informationsverarbeitung des Agenten gehören. In gewisser Weise wird so die scheinbare kategoriale Kluft zwischen Körper und Geist überbrückt. Diese Überbrückung schließt einen Übergang von einer kontinuierlichen, numerischen Beschreibungssprache zu einer diskreten, begrifflichen ein.
- Eine große Vielfalt zeitlicher Phänomene wird sichtbar und kann formal analysiert werden. Der dubiose Begriff "Echtzeit" wird mit empirischer und formaler Substanz gefüllt.
- Die strukturelle Stabilität dynamischer Systeme bedeutet Robustheit gegenüber kleinen, kurzfristigen Störungen. Dies erklärt einen wichtigen Anteil der Robustheit natürlicher Agenten, und verhilft künstlichen Agenten zu einer gewissen Robustheit, wenn sie mit entsprechenden Techniken konstruiert werden.
- Die Kontrollierbarkeit vieler Freiheitsgrade bei der Effektorsteuerung wird durch Ordungsparameter theoretisch erklärt und durch selbstorganisierende Systeme (z.B. Neurokontroller) technisch realisierbar.
- Es wird erklärt, warum Entwicklungsprozesse sich in abwechselnde, relativ stabile Phasen und qualitative Sprünge gliedern. Der empirischen Entwicklungsforschung werden mathematische Techniken in die Hand gegeben, die eine viel feinere (insbesondere zeitliche) Modellierung und damit detailliertere Prädiktionen erlauben als die klassischen statistischen Beschreibungstechniken.
- Evolutionäre oder interaktionistische Methoden ermöglichen im Prinzip, daß informationsverarbeitende Systeme qualitativ neue Konzepte entwickeln und sich an unvorhersehbare Umweltgegebenheiten adaptieren. Dies ist eine Vorbedingung für echte Autonomie.
- Bei der Modellierung von Konzepten werden einige klassische Schwierigkeiten ausgeräumt. Sie betreffen z.B. Variabilität, Kontextsensibilität, Gestalteigenschaften, Feature-Konzept-Unterscheidung, die Natur analoger Repräsentationen, und terminologische Zykel.
- Ein dynamisches Gedächtnismodell erklärt einen schnellen, inhaltsadressierten Zugriff, Vorwärtsinferenzen und andere Phänomen, die grundlegend anders und einfacher als in der herkömmlichen logikorientierten von-Neumann-Rechner-Metaphorik verstanden werden.
- Aufmerksamkeit kann als chaotischer "Suchzustand" analysiert werden.

Neben solchen inhaltlichen Vorzügen würde eine Einbeziehung systemtheoretischer Methoden die KI auch einen wissenschaftshistorischen Bruch kitten. Die KI würde wieder an die kybernetischen Wurzeln anbinden, von denen sie in den 50er Jahren entsproß. Umgekehrt könnten vielleicht logik-orientierte KI-Techniken die Systemtheorie befruchten, nach meiner Meinung vor allem über Methoden der nichtfundierten Mengenlehre. Dieser Gedanke kann hier nicht vertieft werden. Für Norbert Wiener war die Logik die ideale Grundlage der Kybernetik: *"If I were to choose a patron saint for cybernetics* [...], *I should have to choose Leibniz. The philosophy of Leibniz centers about tow closely related concepts — that of a universal symbolism and that of a calculus of reasoning. From these are descended the mathematical notation and the symbolic logic of the present day."* (Wiener 1961[2], p.12).

Hiermit mag es gut sein mit der Lobpreisung systemtheoretischer Methoden für die KI. Es gilt nun auch die Bedenklichkeiten wahrzunehmen. Es gibt eine ganze Reihe innerer (5.1) und äußerer (5.2) Hindernisse, die sich einer Verwendung systemtheoretischer Methoden entgegenstellen.

5.1 Innere Hindernisse für eine Verwendung systemtheoretischer Methoden in der KI

In diesem Abschnitt geht es um Schwierigkeiten, die mit systemtheoretischen Modellierungsansätzen als solchen zusammenhängen, unabhängig vom gegenwärtigen Stand der Kunst.

Systemtheoretische Methoden sind generell **eher für die Analyse als für die Konstruktion** von komplexen informationsverarbeitenden Systemen geeignet. Das hängt sicher mit der Herkunft dieser Methoden aus den Naturwissenschaften zusammen, wo es um das Verständnis, nicht um die Erzeugung von realen Systemen geht. Methoden der Synergetik etwa lassen sich kaum für das Design von Agenten verwenden. Beer (1994) behauptet zwar, systemtheoretische Methoden zum Design seines Neurokontrollers für eine Laufmaschine verwendet zu haben. Bei näherem Hinsehen jedoch reduziert sich seine Designstrategie auf die heuristische Wahl eines Netzwerktyps. Seine Verwendung von genetischen Algorithmen zur Parameteroptimierung kann kaum als Teil des Designs gewertet werden. Die ingenieurswissenschaftliche Systemtheorie und Kontrolltheorie besitzt zwar einen reichen Fundus an Entwurfstechniken. Sie setzen jedoch voraus, daß die einzuhaltenden Arbeitsparameter des zu verwirklichenden Systems bereits bekannt sind. Dies ist spätestens dann unmöglich, wenn das System autonom in einem nichttrivialen Sinn sein soll, weil es dann selbstorganisierend seinen eigenen Entwurf (ganz im Sinne des Autopoieseprinzips) erzeugt. Die Frage nach dem initialen Design solcher Systeme ist aber völlig offen.

Einen Schritt in die Richtung "systemtheoretisch fundiertes Design" stellt die Sprache PDL dar, die in der Arbeitsgruppe von Luc Steels zur Programmierung behavior-basierter Roboter entwickelt wurde (ausführliches Beispiel in Steels 1994d). PDL erlaubt es, ein Kontrollprogramm im Sinner einer diskreten Näherung eines kontinuierlichen dynamischen Systems durch Differenzengleichungen zwischen sog. *quantities* zu spezifizieren. Dies sind Sensorwerte, Effektorsteuerwerte und motivationale Variable. Ein mit PDL geschriebenes Programm kann (wenn keine "hacks" dazwischenkommen) gut als dynamisches System rekonstruiert werden. Das Problem eines systemtheoretisch fundierten Designs ist damit aber noch nicht gelöst, sondern es kann nur besser formuliert werden. Denn es ist bis dato unklar, nach welchen höheren Designprinzipien (z.B. Modularisierung vs. Distribution, explizites Design vs. *on-line evolution*) am besten mit diesem Basiswerkzeug zu verfahren ist. Einen ersten Ansatz zu einem systemtheoretisch fundierten Design mit PDL unternehme ich in Jaeger (1995), wo ich Teilsysteme unter geeigneten Randbedingungen so kopple, daß ihr qualitatives Verhalten trotz Interaktion und Sensordateneinfluß erhalten bleibt. Dies ist ein Beitrag zu einer modularen Designstrategie.

Auch wenn man sich auf die Analyse von intelligenten Systemen beschränkt, sind systemtheoretische Methoden einigen intrinsischen Beschränkungen unterworfen. Hier sollte man wohl zwei Sorten von Schwierigkeiten unterscheiden, nämlich die empirische Analyse physikalisch bereits realisierter Systeme, und die mathematische Analyse von formalen Agentenmodellen.

Bei der **empirischen Analyse** bereits realisierter Systeme geht es darum, aus verrauschtem Datenmaterial ein formales Modell zu konstruieren. Beispiele für solche Analysen sind die Modellierungen von periodischen Handbewegungen von Schöner et al. (1986) oder die Rekonstruktion eines chaotischen Attraktors als Modell der Spontanbewegungen von Neugeborenen (Robertson et al. 1993). Die Vorgehensweise ist in beiden Fällen verschieden. Schöner et al. deuten (in einem kreativen wissenschaftlichen Akt) die empirischen Handbewegungen als das Resultat der Dynamik zweier gekoppelter Oszillatoren. Die für solche Systeme bekannten Gleichungen untersuchen sie dann und stellen eine gute numerische Übereinstimmung mit den empirischen Daten fest. Ihr Vorgehen ist also modellbasiert. Robertson et al. dagegen rekonstruieren "lediglich" einige numerische Kenngrößen des Phasenportraits, um einige prinzipielle, psychologisch relevante Aussagen über Chaos vs. Rauschen und die Dimensionalität des Systems machen zu können.

So verschieden die beiden Arbeiten auch sind, so haben sie doch zweierlei gemeinsam: erstens ist das jeweils untersuchte System sehr einfach, und zweitens gelangen beide Forschungsteams mit den vorhandenen Techniken bereits an die Grenzen des derzeit Möglichen. Man kann vermuten, daß dieses an-Grenzen-Stoßen nicht auf einen ungenügenden Entwicklungsstand der verfügbaren Analysemethoden zu schieben ist, sondern auf eine prinzipielle Erkenntnisgrenze hinweist. Sie besteht darin, daß wirklich komplexe, kontinuierliche dynamische Systeme in ihrem qualitativen Verhalten typischerweise sensibel von einer Vielzahl von Parameterwerten abhängen. Solche Parameter in der erforderlichen Genauigkeit aus empirischen Daten zu rekonstruieren ist unmöglich, wenn die theoretisch erforderliche Information der ausreichend präzisierten Parameter größer ist als die Information, die praktisch aus empirischen, typischerweise sehr verrauschten Daten gewonnen werden kann. Einfacher gesagt: man kann nicht genau genug messen.

Diese Lage ist hinwiederum nicht ganz hoffnungslos. Erstens kann man versuchen, informationsärmere, evtl. diskret approximierende Systemmodelle zu verwenden (eine meiner Motive für die Entwicklung der dynamischen Symbolsysteme, vgl. Jaeger 1994a,d). Zweitens kann man versuchen, das zu analysierende System in Teilsysteme zu zergliedern und diese einzeln zu rekonstruieren. Diese Taktik entspricht dem bewährten Standard der naturwissenschaftlichen Praxis. Allerdings muß man sehen, daß Teilsysteme in informationsverarbeitenden Systemen durch die Kopplungsinteraktion oft in ihrem qualitativen Verhalten verändert werden. Hier macht eine isolierte Untersuchung der Teilsysteme dann wenig Sinn (vgl. Diskussion über schnell variierende Kopplungsparameter in 4.2.4).

Tiefliegende scaling-up-Probleme finden sich auch bei der systemtheoretischen **Analyse formal spezifizierter Systeme**. Hier liegt ein mathematisches Modell des Systems bereits vor, und es geht um dessen qualitative Beschreibung (Attraktoren, Bifurkationen, Stabilitätsbereiche u.a.). Bekanntermaßen sind die Lösungen von Systemen nichtlinearer Differentialgleichungen so gut wie nie analytisch angebbar. Deswegen müssen schon die Untersuchungen solch elementarer Systeme wie die einzelner Taktgeber-Neurone (Nomura et al. 1994) numerische Simulationen durchführen, um ein Bifurkationsdiagramm zu erhalten. In Fällen, die so einfach sind wie dieser, hat man mit solchen Diagrammen ein befriedigendes Bild des Systemverhaltens. Bei komplexeren Systemen kann das Systemverhalten aber auch durch numerische Simulationen nurmehr punktuell für ausgewählte Randbedingungen exploriert werden. Beispiele sind die Untersuchungen des durch Differentialgleichungen modellierten Riechhirns von Yao & Freeman (1990) oder des sensomotorischen Kontrollsystems zur Trajektoriensteuerung bei Gaudiano & Grossberg (1991). Solche Explorationen sind keine geringe Leistung, aber es läuft auf inspiriertes Ausprobieren heraus.

Man kann also festhalten, daß systemtheoretische Beschreibungen des qualitativen Verhaltens empirischer oder formaler Systeme an Komplexitätsgrenzen stoßen. Der faszinierende Anspruch systemtheoretischer Methoden, komplexe Systeme als Ganze verständlich zu machen, kann de facto für intelligente Systeme nicht eingelöst werden. Die in der Physik und der Chemie mit so großem Erfolg untersuchten Systeme sind eben um Größenordnungen weniger komplex als jedes zur Intelligenzerzeugung ansetzende informationsverarbeitende System. Letztere Systeme bestehen typischerweise aus heterogenen Modulen, überspannen mehrere Granularitätsebenen, sind schnell und hochdimensional variierenden Randbedingungen ausgesetzt, wohingegen die "naturwissenschaftlichen" Systeme relativ homogen sind, höchstens zwei Granularitätsebenen umspannen (eine Mikroebene und eine emergente Makroebene), und typischerweise unter niedrigdimensionalen, langsam wechselnden Randbedingungen untersucht werden. Da schon in der Physik systemtheoretische Methoden rasch ausgereizt werden, kann es nicht verwundern, daß sie bei der Analyse intelligenter Systeme überfordert sind.

Es gibt schließlich fundamentale Bedenken gegenüber systemtheoretischen Methoden, die damit zu tun haben, daß mentale Prozesse möglicherweise kategorial verschieden sind von "bloß" physikalischen. Systemtheoretische Techniken sind für erstere entwickelt worden. Aslin (1993) diskutiert einige sich aus dieser Perspektive ergebende Schwierigkeiten.

Generell ist ferner zu bemerken, daß die heute von der Physik und Mathematik gelieferten **mathematischen Methoden keinen stabilen Stand erreicht** haben. Die mathematische und konzeptuelle Entwicklung in der naturwissenschaftlichen und mathematischen Theorie dynamischer Systeme ist in vollem Schwung. Wichtige Grundsatzfragen sind noch nicht richtig verstanden, und die Verwendung systemtheoretischer Methoden geschieht manchmal auf schwankem Grund. Hier sei beispielhaft auf das ungenügende Verständnis des Wechselspiels zwischen Chaos, Noise und der Wahl einer Beschreibungsebene (vgl. Millonas 1994), und Fehlinterpretationenen bei der Rekonstruktion von Attraktoren aus Intervalldaten (Preißl et al. 1990) hingewiesen. Zentrale Begriffe, darunter "Chaos" und "Selbstorganisation" werden auch von Mathematikern ohne präzise Definition gebraucht. Offenbar gibt es verschiedene Mechanismen der Selbstorganisation auf verschiedenen Zeitskalen, aber eine Klassifizierung ist nicht in Sicht. Diese Liste solcher Ungereimtheiten ließe sich verlängern. Für die KI bedeutet das, daß man kein fertiges, in jeder Hinsicht verstandenes Instrumentarium geliefert bekommt.

Aus den verschiedenen in 4.2 referierten Forschungsbeiträgen ist ersichtlich, daß Methoden sowohl der naturwissenschaftlichen Systemtheorie als auch der Kontrolltheorie, und überdies auch evolutionäre Techniken wichtige Beiträge zur Modellierung intelligenter Systeme erbringen. Leider sind die von diesen Forschungstraditionen zur Verfügung gestellten **formalen Techniken derzeit nicht integriert**. Es fehlt einer praktischen Verbindung zwischen den Themenbereichen Regelung/Kontrolle, Selbstorganisation/kollektive Systeme, und langfristiger, evolutionärer Adaptation. In biologischen Agenten sind diese drei Aspekte aber miteinander verschränkt. Hier helfen uns systemtheoretische Denkweisen gegenwärtig noch nicht viel weiter.

5.2 Äußere Hindernisse für eine Verwendung systemtheoretischer Methoden in der KI

Erhebliche Hürden für die Verwendung systemtheoretischer Methoden in der KI resultieren aus "äußeren", soziokulturellen Bedingungen:

- Formale systemtheoretische Kenntnisse sind bei KI-Forschern kaum vorhanden. Da gegenwärtige systemtheoretische Lehrbücher einen gediegenen mathematischen, physikalischen oder ingenieurswissenschaftlichen Hintergrund des Lernenden voraussetzen, dürfte das Erlernen systemtheoretischer Methoden über Gebühr Schwierigkeiten bereiten und demzufolge nicht erfolgen.
- Die Verwendung von logik-orientierten Techniken macht inzwischen einen guten Teil des Selbstbildes der KI-Forschungsgemeinde aus. Hieraus resultiert wohl bei Vielen eine reflexhafte Abwehr systemtheoretischer Techniken.
- Gegenwärtig werden dynamische Systeme als Modell situierter Agenten oder als neues Paradigma für die Kognitionswissenschaft von einigen Wenigen in euphorischen Tönen empfohlen. Die in 5.1 beschriebenen Schwierigkeiten des neuen, alten "Paradigmas" garantieren aber, daß hier geweckte Erwartungen nicht rasch befriedigt werden können. Es besteht also die Gefahr, daß — wie so oft in der Geschichte der KI — nach einer kurzen Euphorie eine so starke Ernüchterung erfolgt, daß auch langfristig produktive Ansätze wegen versiegender Mittel wieder aufgegeben werden.
- Die systemtheoretische Perspektive liefert nach meiner Meinung vorerst mehr für die Analyse von intelligenten Systemen als für deren Design. Das steht überkreuz mit der gegenwärtigen Anwendungsorientierung in der KI.

6 Ausblick

Diese Arbeit hat drei Ziele verfolgt. Erstens sollte sie dem Leser einen Überblick über bestehende Systemtheorien und einige wichtige systemtheoretische Konstrukte vermitteln. Zweitens wollte sie ein Gefühl dafür vermitteln, welche enormen Erklärungspotentiale systemtheoretische Modellbildungen für die KI haben könnten. Und drittens wollte sie ebenfalls ein Gefühl für die nicht zu unterschätzenden Schwierigkeiten vermitteln, welche ebendiese Modellbildungen mit sich bringen.

Ich hoffe, daß beim Leser am Ende die Antizipation der Potentiale stärker ist als die Angst vor den Schwierigkeiten. Oder besser noch, daß die Schwierigkeiten als eine Herausforderung zur Erschließung neuer Horizonte für unser Verständnis intelligenter Systeme empfunden werden. Wer bei dieser Haltung angelangt ist, der wird sich fragen: wie soll es weitergehen? bzw., wie soll es jetzt anfangen?

Die Schwierigkeiten bei der Verwendung systemtheoretischer Methoden wurzeln zu einem guten Teil darin, daß diese Methoden aus Gebieten stammen (Physik, Naturwissenschaften, Ingenieurswissenschaften), in denen weniger komplexe Systeme behandelt werden als in einer ernstzunehmenden KI. Für die KI bedeutet das, daß sie ihre eigenen Beiträge zu einer Weiterentwicklung systemtheoretischer Methoden leisten muß. Mir fallen auf Anhieb die folgenden Fragenkomplexe ein:

- die Beziehungen zwischen (1) spontaner Adaptation im Sinne von situierter Reflexivität, (2) struktureller Differenzierung im Sinne von selbstorganisierendem Lernen, und (3) Systemoptimierung in evolutionärer Perspektive,
- eine Observablentheorie, die eine Abbildung von empirischen, operational identifizierbaren Entitäten (z.B. Konzepte, Behaviors, mentale Zustände) auf formale systemtheoretische Konstrukte (eben nicht einfach Attraktoren oder Bereiche im Phasenraum) leistet,
- Zeitskalen, die selbst dynamisch veränderlich sind (d.h. keine feste Unterscheidung Kontrollparameter vs. Zustandsvariable),
- hierarchische Systeme mit mehreren (mehr als 2) emergenten Ebenen, wobei Entwicklungsprozesse diese aus einem primären System mit nur einer Ebene entstehen lassen,
- eine informationstheoretische, interaktionistische Klärung des Begriffs der Repräsentation,
- die Beziehungen zwischen seriell-zeitlichen vs. parallel-räumlichen Organisationsprinzipien.

All dies sind anspruchsvolle Themen, denen sich auch die bekannten Systemtheorien schon — ohne durchgreifenden Erfolg — gewidmet haben. Es erscheint darum aussichtslos, bekannte Methoden einfach nur auszuweiten oder verbessern zu wollen. Die KI sollte wohl eigene formale Systemtypen entwickeln, die in geeigneter Weise einfacher oder gröber sind als kontinuierliche dynamische Systeme. Man darf gespannt sein — schließlich ist die KI doch eine dynamische Wissenschaft!

Danksagung

Die Motivation zu dieser Arbeit verdanke ich langen Diskussionen über dynamische Systeme und autonome Agenten, die ich mit Simone Strippgen und Thomas Christaller geführt habe. Luc Steels und seine Mitarbeiter Peter Stuer, Bart de Boer, sowie Emmet Spier haben durch ihr anhaltendes Interesse und motivierenden Zuspruch der Sache zum Fortgang und glücklichen Ende verholfen. Wichtige Hinweise verdanke ich Gregoire Nicolis, Donal McKernan, und Robert Port. Ihnen allen möchte ich herzlich danken.

Die Arbeit wurde in einer ersten Version durch einen Werkvertrag von der GMD in St. Augustin finanziert; wesentliche Ergänzungen konnte ich dank eines Forschungsstipendiums der VUB Brüssel vornehmen.

Literatur

Abraham, Ralph H., Shaw, Christopher D. (1983f): Dynamics: The Geometry of Behavior. Parts 1-4. Aerial Press, Santa Cruz, California. Nachdruck bei Addison-Wesley, Redwood City 1992

Aczel, P. (1988): Non-Well-Founded Sets. CSLI Lecture Notes 14, CSLI, Menlo Park, Kalifornien

Arrowsmith, D.K., Place, C.M. (1992): Dynamic Systems - Differential Equations, Maps and Chaotic Behavior. Chapman and Hall, London 1992

Arrowsmith, D.K., Place, C.M. (1990): An Introduction to Dynamical Systems. Cambridge University Press, Cambridge 1990. Deutsch: Dynamische Systeme. Spektrum Akademischer Verlag, Heidelberg/Berlin/Oxford 1994

Aslin, Richard N. (1993): The Strange Attractiveness of Dynamic Systems to Development. In Smith & Thelen 1993a, 358-399

Baas, N.A. (1994): Emergence, Hierarchies, and Hyperstructure. In: Langton, C.G. (Hrsg.) (1994): Artificial Life III. SFI Studies in the Sciences of Complexity, Proc. Vol. XVII, Addison-Wesley, 515-537

Baasye, K., Dean, Th., Kaelbling, L.P. (1995): Learning Dynamics: System Identification for Perceptually Challenged Agents. Artificial Intelligence 72, 139-171

Babloyantz, A., Lourenço, C. (1994): Computation with Chaos: A Paradigm for Cortical Activity. Proceedings of the National Academy of Sciences of the USA 91, 9027-9031

Barsalou, L.W. (1987): The Instability of Graded Structure: Implications for the Nature of Concepts. In: Neisser, U. (Hrsg.): Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization. Cambridge University Press, Cambridge 1987

Barsalou, L.W. (1989): Intraconcept Similarity and its Implications for Interconcept Similarity. In: Vosniadou, S., Ortony, A. (Hrsg.): Similarity and Analogical Reasoning. Cambridge University Press, 76-121

Bartlett, F.C. (1932): Remembering: A Study in Experimental and Social Psychology. Cambridge University Press, Cambridge 1932 (letzte Neuauflage 1977)

Barwise, J., Moss, L. (1991): Hypersets. The Mathematical Intelligencer 13 (4), 1991, 31-41

Beer, R. (1995): A Dynamical Systems Perspective on Agent-Environment Interaction. Artificial Intelligence 72 (1/2), 173-216

Bennett, C.H. (1991): The Thermodynamics of Computation: a Review. International Journal of Theoretical Physics 21 (12), 1982, 905-940

von Bertalanffy, Ludwig (1968): General System Theory: Foundations, Development, Applications. Brazillier, New York

Bertenthal, B.I., Pinto, J. (1993): Complementary Processes in the Perception and Production of Human Movements. In: Smith & Thelen 1993a, 209-239

Bickhard, M. (1993): Representational Content in Humans and Machines. Journal of Experimental and Theoretical AI 5 (1993), 285-333

Blumberg, B. (1994): Action Selection in Hamsterdam. In: Cliff, D. et al. (Hrsg.) (1994): From Animals to Animats III. Proceedings der dritten internationalen Konferenz "On the Simulation of Adaptive Behavior", Bradford/MIT Press, 108-117

Bookman, L.A. (1988): A Connectionist Scheme for Modelling Context. In: Touretzky, D., Hinton, G., Sejnowski, T. (Hrsg.): Proceedings of the 1988 Summer School on Connectionist Models, Morgan Kaufmann 1988, 281-290

Bookman, L.A., Sun, R. (1993) (Hrsg.): Architectures for Integrating Neural and Symbolic Processes. Special Issue of Connection Science, Vol. 5 (3&4)

Braitenberg, V. (1978): Cell Assemblies in the Cerebral Cortex. In: Heim, R., Palm, G. (Hrsg.) (1978): Theoretical Approaches to Complex Systems. Springer Verlag, Berlin/Heidelberg/New York (Lecture Notes in Biomathematics 21), 171-188

Brézillon, P. (1993) (Hrsg.): Proceedings des IJCAI-93 Workshop "Using Knowledge in its Context". Technical Report LAFORIA 93/13, Institut Blaise Pascal, Université Paris VI et VII

Briggs, J., Peat, D. (1989): Turbulent Mirror. Harper & Row, New York

Brooks, R.A. (1989): The Whole Iguana. In: Brady, M. (Hrsg.): Robotics Science. MIT Press, Cambridge, Mass., 1989, 432-456

Brooks, R.A. (1991): New Approaches to Robotics. Science 253, 1227-1232

Buhmann, J., Schulten, K. (1987): Storing Sequences of Biased Patterns in Neural Networks with Stochastic Dynamics. In: Eckmiller, R., von der Malsburg, C. (Hrsg.) (1987): Neural Computers. Proceedings des NATO Advanced Research Workshop on Neural Computers 1987. Springer Verlag, Berlin/Heidelberg/New York 1987, 231-242

Cabanac, M. (1992): Pleasure: The Common Currency. Journal of theoretical Biology 155 (1992), 173-200

Caelli, T.M., Squire, D.McG., Wild, T.P.J. (1993): Model-Based Neural Networks. Neural Networks 6 (5), 613-625

Campbell, G.J., Taylor, J.G. (1993): The Temporal Kohonen Map. Neural Networks 6, 441-445

Carlyle, J.W. (1969): Stochastic Finite-state System Theory. In Zadeh & Polak 1969, 387-422

Carpenter, G.A., Grossberg, S. (1990): ART 3: Hierarchical Search Using Chemical Transmitters in Self-Organizing Pattern Recognition Architectures. Neural Networks 3 (2), 129-152

Casdagli, M. (1992): A Dynamical Systems Approach to Modeling Input-Output Systems. In: Casdagli, M., Eubank, S. (Hrsg.) (1992): Nonliner Modeling and Forecasting. SFI Studies in the Sciences of Complexity, Proc. Vol. XII, Addison-Wesley

Chalmers, D.J. (1990): Subsymbolic Computation and the Chinese Room. In: Dinsmore, J. (Hrsg.): The Symbolic and Connectionist Paradigms: Closing the Gap. Lawrence Erlbaum, Hillsdale N.J., 1992, 25-48

Chalmers, D.J. (1992): Connectionism and Compositionality: Why Fodor and Pylyshyn Were Wrong. Philosophical Psychology 6, 1993, 305-319

Clancey, W.J. (1993): Situated Action: A Neuropsychological Interpretation. Cognitive Science 17, No 1, 1993, 87-116

Clark, A. (1994): Autonomous Agents and Real-Time Success: Some Foundational Issues. Eingereicht bei der IJCAI-95, Vorabdruck in den Proceedings des Workshops "On the Role of Dynamics and Represenation in Adaptive Behavior and Cognition", Universidad del Pais Vasco, San Sebastian 1994, 19-22

Clark, J.E., Truly, T.L., Phillips, S.J. (1993): On the Development of Walking as a Limit-Cycle System. In: Smith & Thelen 1993a, 71-94

Clutton-Brock, T.H., Harvey, P.H. (Hrsg.) (1978): Readings in Sociobiology. Freeman, Reading/San Francisco

Cruse, H., Bartling, Ch., Cymbalyuk, G., Dean, J., Dreifert, M. (1994): A Modular Artificial Neural Net for Controlling a Six-Legged Walking System. Proceedings der Konferenz "Prerational Intelligence in Robotics: from Sensorimotor Intelligence to Collective Behavior", Report Nr. 10 der Forschungsgruppe "Prerational Intelligence", Zentrum für interdisziplinäre Forschung (ZIF), Universität Bielefeld, 17-36

Crutchfield, J.P., Young, K. (1990): Computation at the Onset of Chaos. In: Zurek, W.H. (Hrsg.): Complexity, Entropy, and the Physics of Information. SFI Studies in the Sciences of Complexity, vol. VIII, Addison-Wesley, 1990, 223-269

Crutchfield, J.P. (1992): Semantics and Thermodynamics. In: Casdagli, M., Eubank, S. (eds.) (1992): Nonlinear Modeling and Forecasting. SFI Proceedings Vol. XII, Addison-Wesley, Redwood City etc., 317-359

Dalenoort, G.J. (1990a): Towards a General Theory of Representation. Report 31/1990 der Forschungsgruppe "Mind and Brain", Zentrum für interdisziplinäre Forschung (ZIF), Universität Bielefeld

Dalenoort, G.J. (1990b): Self-Organization of Informational Systems. Report 47/1990 der Forschungsgruppe "Mind and Brain", Zentrum für interdisziplinäre Forschung (ZIF), Universität Bielefeld

Dautenhahn, K., Christaller, Th. (1995): Remembering, Rehearsal and Empathy: Inspirations from a Dynamical Systems Point of View. Erscheint in den Proceedings der "Reaching for Mind" Workshop auf der AISB-95

Davies, Paul (1987): Cosmic Blueprint. Deutsch: Prinzip Chaos - die neue Ordnung des Kosmos. Bertelsmann, München 1988

Deneubourg, J.L. (1977): Application de l'ordre par fluctuations a la description de certaines étapes de la construction du nid chez les termites. Insectes Sociaux 24 (2), 117-130

Deneubourg, J.L, Goss, S., Franks, N., Sendova-Franks, A., Detrain, C., Chrétien, L. (1991): The Dynamics of Collective Sorting: Robot-Like Ants and Ant-Like Robots. In: Meyer, J.A., Wilson, S. (Hrsg.) (1991): From Animals to Animats 1. Proceedings der First International Conference on the Simulation of Adaptive Behavior. MIT Press, 356-365

Deneubourg, J.L., Franks, N.R. (1995): Collective Control without Explicit Coding: The Case of Communal Nest Excavation. Technical Report ULB-Cenoli 95-10, Centre for Nonlinear Phenomena and Complex Studies, Université Libre de Bruxelles (erscheint in J. of Insect Behavior)

Dorffner, G. (1994): Repräsentation und Selbstorganisation im Konnektionismus. In: Duwe, I., Kurfeß, F., Paaß, G., Vogel, S. (Hrsg.) (1994): Konnektionismus und Neuronale Netze. Beiträge zur HeKoNN94. GMD-Studien 242, Gesellschaft für Mathematik und Datenverarbeitung (GMD), St. Augustin, 213-228

Drescher, G.L. (1991): Made-up Minds: A Constructivist Approach to Artificial Intelligence. MIT Press, Cambridge, Mass.

Dress, A., Hendrichs, H., Küppers, G. (Hrsg.)(1986): Selbstorganisation - die Entstehung von Ordnung in Natur und Gesellschaft. Piper, München/Zürich 1986

Ebeling, W., Nicolis, G. (1992): Word Frequency and Entropy of Symbolic Sequences: a Dynamical Perspective. Chaos, Solitons & Fractals 2 (6), 635-650

Eckhorn, R., Stoecker, M. (1994): Synthesizing Complex Perceptions I: Concepts of Visual Feature Associations Based on Neural Synchronization and Related Experimental Results in the Visual Cortex. Proceedings der Konferenz "Emergence of Prerational Intelligence in Biology: From Sensorimotor

Intelligence to Collective Behavior", Part 1, Report Nr. 7 der Forschungsgruppe "Prerational Intelligence", Zentrum für interdisziplinäre Forschung (ZIF), Universität Bielefeld, 81-100

Eckmiller, R. (1987): Neural Networks for Motor Program Generation. In: Eckmiller, R., von der Malsburg, C. (Hrsg.) (1987): Neural Computers. Springer Verlag, Berlin/Heidelberg/New York, 361-371

von Ehrenfels, Ch. (1890): Über "Gestaltqualitäten". Vierteljahresschrift für wissenschaftliche Philosophie XIV 3, 1890. Nachdruck in: Weinhandl, F. (Hrsg.): Gestalthaftes Sehen. Wissenschaftliche Buchgesellschaft, Darmstadt 1978, 11-43

Eigen, M., Schuster, P. (1977f): The Hypercycle: A Principle of Natural Self-Organization. Naturwissenschaften 64 (1977), 541-565 (Part A); 65 (1978), 7-41 (Part B); 65 (1978), 341-369 (Part C)

Eigen, M., Winkler, R. (1975): Das Spiel. Piper, München

Elman, J.L. (1990): Finding Structure in Time. Cognitive Science 14 Nr. 2, 179-211

Engel, A.K:, König, P., Gray, C.M., Singer, W. (1990): Synchronization of Oscillatory Responses: A Mechanism for Stimulus-Dependent Assembly Formation in Cat Visual Cortex. In: Eckmiller, R, Hartmann, G., Hauske, G. (Hrsg.) (1990): Parallel Processing in Neural Systems and Computers. Elsevier/North Holland, 105-108

Engeler, E. (1983): Metamathematik der Elementarmathematik. Springer Verlag, Berlin/Heidelberg/New York

Engels, Ch., Schöner, G. (1995): Dynamic Fields Endow Behavior-Based Robots with Representations. Robotics and Autonomous Systems 14 (1995), 55-77

Ewert, J.-P., von Seelen, W. (1974): Neurobiologie und System-Theorie eines visuellen Mustererkennungsmechanismus bei Kröten. Kybernetik 14 (1974), 167-183

Flavell, J.H. (1968): The Developmental Psychology of Jean Piaget. Van Nostrand, Princeton 1968

Fodor, J.A., Pylyshin, Z. (1988): Connectionism and Cognitive Architecture: A Critical Analysis. Cognition 28, 3-71

von Foerster, H. (1984a): Erkenntnistheorien und Selbstorganisation. DELFIN 1984 (4), 6-19. Nachdruck in Schmidt, S.J. 1987a, 133-158

von Foerster, H. (1984b): Principles of Self-Organization in a Socio-Managerial Context. In: Ulrich, H., Probst, G.J.B. 1984, 2-24

Ford, Joseph (1989): What is Chaos, that We Should be Mindful of It? In: Davies, Paul (Hrsg.) (1989): The New Physics. Cambridge University Press, Cambridge, p. 348-371

Forrest, S. (Hrsg.) (1990): Emergent Computation: Self-organizing, Collective, and Cooperative Phenomena in Natural and Artificial Computing Networks. Proceedings of the 9th annual CNLS conference, Los Alamos 1989. Physica D 42

Forrest, S., Miller, J.H. (1990): Emergent Bhavior in Classifier Systems. Physica D 42 (1990), 213-227

Forrester, J.W. (1972): Principles of Systems. Deutsch: Grundzüge einer Systemtheorie. Betriebswirtschaftlicher Verlag Dr. Th. Gabler, Wiesbaden

Gaudiano, P., Grossberg, S. (1991): Vector Associative Maps: Unsupervised Real-Time Error-Based Learning and Control of Movement Trajectories. Neural Networks 4 (2), 1991, 147-183

van Geert, P. (1993): A Dynamic Systems Model of Cognitive Growth: Competition and Support Under Limited Resource Conditions. In: Smith & Thelen 1993a, 265-332

van Gelder, T., Port, R. (1994a): Beyond Symbolic: Toward a Kama-Sutra of Compositionality. In: Honavar, V., Uhr, L. (Hrsg.): Artificial Intelligence and Neural Networks: Steps Toward Principled Integration. Academic Press, 1994, 107-125

van Gelder, T., Port, R. (1994b): It's About Time: An Overview of the Dynamical Approach to Cognition. Research Report 116, Cognitive Science Programme, Indiana University, Bloomington, Indiana. Erscheint auch als Vorwort zu van Gelder & Port (1995)

van Gelder, T., Port, R. (1995) (Hrsg.): Mind as Motion: Explorations in the Dynamics of Cognition. Bradford/MIT Press, erscheint voraussichtlich im Frühjahr 1995

Georgeff, M.P., Lansky, A.L. (Hrsg.)(1986): Reasoning about Actions and Plans. Proc. eines Workshops in Timberline, Oregon 1986. Morgan Kaufmann, Los Altos

Gerhardt, M., Schuster, H. (1989): A Cellular Automaton Describing the Formation of Spatially Ordered Structures in Chemical Systems. Physica D 36, 209-221

Gill, A. (1969): Finite-state Systems. In Zadeh & Polak 1969, 43-94

Goldberg, D.E. (1989): Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading, Mass. 1989

Goschke, T., Koppelberg, D. (1990): Connectionist Representation, Semantic Compositionality, and the Instability of Concept Structure. Psychological Research 52, 253-270

Gould, S.J. (1979): Ever Since Darwin. Reflections in Natural History. Norton, New York/London

Greenfield, P. (1991): Language, Tools and Brain: The Ontogeny and Phylogeny of Hierarchically Organized Sequential Behavior. Behavioral and Brain Sciences 14, 531-595

Greeno, J.G., Moore, J.L. (1993): Situativity and Symbols: Response to Vera and Simon. Cognitive Science 17, No 1, 1993, 49-60

Grossberg, S., Somers, D. (1991): Synchronized Oscillations During Cooperative Feature Linking in a Cortical Model of Visual Perception. Neural Networks 4 (4), 1991, 453-466

Haken, H. (1983): Advanced Synergetics - Instability Hierarchies of Self-Organizing Systems and Devices. Springer, Berlin/Heidelberg (Springer Series in Synergetics Vol. 20)

Haken, H., Wunderlich, A. (1986): Synergetik: Prozesse der Selbstorganisation in der belebten und unbelebten Natur. In Dress et al. 1986, 35-60

Happel, B.L.M., Murre, J.M.J. (1995): Evolving Complex Dynamics in Modular Interactive Neural Networks. Preprint, Universität Leiden, eingereicht bei Neural Networks

Harnad, S. (Hrsg.) (1987): Categorical Perception. Cambridge University Press, Cambridge, Mass.

Hart, J., Berndt, R.S., Caramazza, A. (1985): Category-specific Naming Deficit Following Cerebral Infarction. Nature 316, 439-440

Harvey, I., Husbands, P., Cliff, D. (1994): Seeing the Light: Artificial Evolution, Real Vision. In: Cliff, D. et al. (Hrsg.) (1994): From Animals to Animats III: Proc. der Third Int. Conf. on Simulation of Adaptive Behavior. Bradford/MIT Press, 392-401

Hasida, K. (1994): Dynamics of Symbol Systems. New Generation Computing 12 (1994), 285-310

Hayashi, Y. (1994): Oscillatory Neural Networks and Learning of Continuously Transformed Patterns. Neural Networks 7 (2), 219-232

Hayes, P.J. (1985): Some Problems and Non-Problems in Representation Theory. In: Brachman, R.J., Levesque, H.J. (Hrsg.) (1985): Readings in Knowledge Representation. Morgan Kaufman, Los Altos, 4-22

Hayes, P.J., Ford, K.M., Agnew, N. (1994): On Babies and Bathwater. A Cautionary Tale. AI Magazine, Fall 1994, 15-26

Hebb, D.O. (1949): The Organization of Behavior. Wiley, New York

Hofstadter, D.R. (1979): Gödel, Escher, Bach. Basic Books, New York

Hofstadter, D.R., Mitchell, M. (1993): The Copycat Project: A Model of Mental Fluidity and Analogy-Making. Erscheint in: Holyoak, K., Barnden, J. (Hrsg.): Advances in Connectionist and Neural Computation Theory, Vol. II: Analogical Connections. Ablex, Norwood, N.J. Leicht überarbeitet auch als Kap. 5 in Hofstadter, D. (1995): Fluid Concepts and Creative Analogies. Basic Books, New York

Holland, J.H. (1975): Adaptation in Natural and Artificial Systems. The University of Michigan Press, Ann Arbor, 1975

Holland, J.H. (1986): Escaping Brittleness: The Possibilities of General-Purpose Learning Algorithms Applied to Parallel Rule-Based Systems. In: Anderson, J.R., Michalski, R.S., Carbonell, J.G. (Hrsg.) (1986): Machine Learning. Kaufmann, Los Altos, 593-623

Holland, J.H., Reitmann, J.S. (1978): Cognitive Systems Based on Adaptive Algorithms. In: Waterman, D.A., Hayes-Roth, F. (Hrsg.) (1978): Pattern Directed Inference Systems. Academic Press, New York, 313-329

Jaeger, H. (1991): An Introduction to Dynamic Concept Systems. In: Boley, H., Richter, M.M. (Hrsg.), Processing Declarative Knowledge. Proceedings of the PDK-91 at Kaiserslautern, Springer Verlag, Berlin (Lecture Notes in Artificial Intelligence 567), 87-106

Jaeger, H. (1992): A Type-free Semantics for Concepts in Contexts. In: Brézillon 1993, 51-61

Jaeger, H. (1994a): Dynamic Symbol Systems. Dissertation, Technische Fakultät der Universität Bielefeld 1994

Jaeger, H. (1994b): Gestaltbildung durch Informationsmaximierung: eine formale und algorithmische Rekonstruktion. In: Opwis, K. (Hrsg.) (1994): Proceedings "Erste Fachtagung der deutschen kognitionswissenschaftlichen Gesellschaft", Psychologisches Institut, Universität Freiburg 1994, 32-34

Jaeger, H. (1994c): On Modelling Behaviors and Concepts as Attractors. Proceedings der Konferenz "On the Role of Dynamics and Representation in Adaptive Behaviour and Cognition (DRABC-94)" in San Sebastian, Dezember 1994. Universidad del País Vasco, San Sebastian 1994, 171-173

Jaeger, H. (1994d): An Introduction to Dynamic Symbol Systems. In: Hallam, J. (Hrsg.): Hybrid Problems, Hybrid Solutions. Proceedings van de AISB-95. IOS Press/Ohmsha, Amsterdam 1995, 109-120

Jaeger, H. (1995): Modulated Modules: Coupling Subsystems While Preserving Structural Stability. Erscheint als GMD-Report

Kaelbling, L.P. (1991): Foundations of Learning in Autonomous Agents. Robotics and Autonomous Systems 8 (1-2), 1991. Nachdruck in: Van der Velde, W. (Hrsg.) (1993): Toward Learning Robots. Bradford/MIT Press

Kawato, M. (1994): A Bi-Directional Theory Approach to Prerational Intelligence. In: Proceedings der Konferenz "Prerational Intelligence in Robotics: from Sensorimotor Intelligence to Collective Behavior", Report Nr. 10 der Forschungsgruppe "Prerational Intelligence", Zentrum für interdisziplinäre Forschung (ZIF), Universität Bielefeld, 71-83

Kelso, J.A. Scott, Ding, Mingzhou, Schöner, Gregor (1993): Dynamic Pattern Formation: A Primer. In: Smith, Linda B., Thelen, Esther (Hrsg.) (1993): A Dynamic Systems Approach to Development. Bradford/MIT Press, Cambridge, Mass. 1993, 13-50

Kiss, G. (1993): Autonomous Agents, AI and Chaos Theory. In: Meyer, J.A., Roitblatt, H.L., Wilson, S.W. (Hrsg.): From Animals to Animats 2. Proceedings of the Second Int. Conf. on Simulation of Adaptive Behavior, 518-524

Köhler, R. (1987): System Theoretic Linguistics. Theoretical Linguistics 14 (1987), 241-257

Kornwachs, K. (Hrsg.) (1984): Offenheit - Zeitlichkeit - Komplexität: Zur Theorie der offenen Systeme. Campus, Frankfurt/New York

Košecká, J., Bajcsy, R. (1994): Discrete Event Systems for Autonomous Mobil Agents. Robotics & Autonomous Systems 12 (1994), 187-198

Krause, W. (1989): Über menschliches Denken - Denken als Ordungsbildung. Zeitschrift für Psychologie 197, 1989, 1-30

Kriz, Jürgen (1992): Chaos und Struktur: Systemtheorie Band 1. Quintessenz Verlag, München

Krohn, W., Küppers, G., Paslack, R. (1987): Selbstorganisation - Zur Genese und Entwicklung einer wissenschaftlichen Revolution. In S.J. Schmidt 1987a, 441-465

Kuperstein, M. (1991): INFANT Neural Controller for Adaptive Sensory-Motor Coordination. Neural Networks 4 (2), 131-145

Lakoff, G. (1987): Women, Fire, and Dangerous Things. The University of Chicago Press, Chicago

Large, E.W., Kolen, J.F. (1995): Resonance and the Perception of Musical Meter. Connection Science 6(1), 177-208

Lashley, K.S. (1951): The Problem of Serial Order in Behavior. In: Jefress, L.A. (Hrsg.) (1951): Cerebral Mechanisms in Behavior, Wiley, 1951, 112-136

van Leeuven, C. (1990): Perceptual Learning Structures as Conservative Structures: Is Economy an Attractor? Technical Report 45/1990 of the Research Group "Mind and Brain" at the Center for Interdisciplinary Research (ZIF), University of Bielefeld 1990

Legendre, G., Miyata, Y., Smolensky. P. (1990a): Harmonic Grammar - A formal multi-level connectionist theory of linguistic well-formedness: An application. ICS Technical Report #90-4, University of Colorado at Boulder

Legendre, G., Miyata, Y., Smolensky. P. (1990b): Harmonic Grammar - A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. ICS Technical Report #90-5, University of Colorado at Boulder

Leshno, M., Lin, V.Y., Pinkus, A., Schocken, S. (1993): Multilayer Feedforward Networks With a Nonpolynomial Activation Function Can Approximate Any Function. Neural Networks 6 (6), 861-867

Ling, C.X., Buchal, R. (1993): Learning to Control Dynamic Systems with Automated Quantization. In: Brazdil, P.B. (Hrsg.): Machine Learning. Proceedings der ECML-93. Springer Verlag, Berlin etc. (Lecture Notes in Artificial Intelligence), 372-377

MacLennan, B. (1990): Synthetic Ethology: An Approach to the Study of Communication. In: Langton, C.G., Taylor, C., Farmer, J.D., Rasmussen, S. (Hrsg.) (1990): Artificial Life II. Addison Wesley, Redwood City etc., 631-658

MacLennan, B. (1991): Continuous Symbol Systems: The Logic of Connectionism. Technical Report CS-91-145, University of Tennesse, Knoxville

Mandelbrot, B.B. (1977): The Fractal Geometry of Nature. Freeman, New York 1977. Deutsch: Die fraktale Geometrie der Natur. Birkhäuser, Basel/Boston 1987

Mandelbrot, B.B. (1986): Fractals and the Rebirth of Iteration Theory. In: Peitgen, H.-O., Richter, P.H. (1986): The Beauty of Fractals. Springer, Berlin/Heidelberg, 151-160

Mangold-Allwinn, R. (1991): Flexible Konzepte: Modelle, Experimente, Simulationen. Habilitationsschrift, University of Mannheim, 1991

Mani, D.R., Shastri, L. (1993): Reflexive Reasoning with Multiple Instantiation in a Connectionist Reasoning System with a Type Hierarchy. Connection Science 5, No 3/4 (1993), 205-242

Matsuga, M., Yuille, A.L. (1994): Spatiotemporal Information Storage in a Content Adressable Memory Using Realistic Neurons. Neural Networks 7 (3), 1994, 419-440

Maturana, H.R., Varela, F.J. (1984): El árbol del concocimiento. Deutsch: Der Baum der Erkenntnis: die biologischen Wurzeln des menschlichen Erkennens. Scherz, Bern/München 1987; Goldmann Verlag, 1992[4]

McFarland, D. (1994): Towards Robot Cooperation. In: Cliff, D. et al. (eds.) (1994): From Animals to Animats III: Proceedings or the Third Int. Conf. on Simulation of Adaptive Behavior. MIT Press/Bradford Books, Cambridge, Mass., 440-444

Medin, D.L. (1989): Concepts and Conceptual Structure. American Psychologist 44 (12), 1469-1481

Medin, D.L., Barsalou, L.W. (1987): Categorization Processes and Categorical Perception. In Harnad 1987, 455-490

Mehl, S. (1992): Dynamische Semantische Netze: Zur Kontextabhängigkeit von Wortbedeutungen. Dissertation, Universität Koblenz-Landau in Koblenz, 1992

Millonas, M. (1994): The Importance of Being Noisy. The Bulletin of the Santa Fe Institute, Summer 1994, 22-23

Moore, C. (1993): Real-valued, Continuous-time Computers: A Model of Analog Computation, Part 1. Preprint, Santa Fé Institute 1993

Müller, J. (Hrsg.) (1993): Verteilte Künstliche Intelligenz. Methoden und Anwendungen. BI Wissenschafts-verlag, Mannheim etc.

Munro, P.W., Anderson, J.A. (1988): Tools for Connectionist Modeling: The Dynamical Systems Methodology. Behavior Research Methods, Instruments, & Computers 20(2), 1988, 276-281

Murphy, G.L., Medin, D.L. (1985): The Role of Theories in Conceptual Coherence. Psychological Review 92 (3), 289-316

Nebel, B. (1990): Reasoning and Revision in Hybrid Representation Systems. Lecture Notes in Artificial Intelligence 422, Springer Verlag, Berlin/Heidelberg/New York

Nebel, B. (1991): Terminological Cycles: Semantics and Computational Properties. In: Sowa, J. (Hrsg.) (1991). Principles of Semantic Networks. Morgan Kaufman, San Mateo, 331-361

Newtson, D. (1993): The Dynamics of Action and Interaction. In Smith & Thelen 1993a, 241-264

Nicolis, J.S., Tsuda, I. (1989): On the Parallel Between Zipf's Law and 1/f Processes in Chaotic Systems Possessing Coexisting Attractors. Progress of Theoretical Physics 82 (2), 254-274

Nicolis, G. (1989): Physics of Far-from-equilibrium Systems and Self-Organization. In: Davies, Paul (Hrsg.) (1989): The New Physics. Cambridge University Press, Cambridge, p. 316-347

Nicolis, G., Nicolis, C., McKernan, D. (1993): Stochastic Resonance in Chaotic Dynamics. Journal of Statistical Physics 70 (1/2), 125-139

Nomura, T., Sato, S., Doi, S., Segundo, J.P., Stiber, M.D. (1994): A Modified Isochron Clock with Slow and Fast Dynamics as a Model of Pacemaker Neurons. Biological Cybernetics 72, 93-101

Nyman, A. (1966): Die Schulen der neueren Psychologie. Verlag Hans Huber, Bern/Stuttgart 1966

Obermayer, K., Ritter, H., Schulten, K. (1990a): A Principle for the Formation of the Spatial Structure of Cortical Feature Maps. Proceedings of the National Academy of Sciences of the USA 87, 1990, 8345-8349

Obermayer, K., Ritter, H., Schulten, K. (1990b): Large-Scale Simulation of a Self-Organizing Neural Network: Formation of a Somatotopic Map. In: Eckmiller, R., Hartmann, G., Hauske, G. (Hrsg.) (1990): Parallel Processing in Neural Systems and Computers. North-Holland, Amsterdam 1990, 71-74

Packard, N., Wolfram, S. (1985): Two-dimensional Cellular Automata. Journal of Statistical Physics 38 (5/6), 901-946. Nachdruck in Wolfram 1986, 126-171

Palm, G. (1988): On the Asymptotic Information Storage Capacity of Neural Networks. In: Eckmiller, R., von der Malsburg, Ch. (Hrsg.) (1988): Neural Computers. NATO ASI Series F 41, Springer Verlag, Berlin/Heidelberg 271-280

Paslack, R. (1989): »...da stellt ein Wort zur rechten Zeit sich ein«. Die Karriere des Chaos zum Schlüssel-begriff. Kursbuch 98, 121-140

Patel, M.J., und Schnepf, U. (1991): Concept Formation as Emergent Phenomena. In: Varela, J.V., Bourgine, P. (Hrsg.): Toward a Practice of Autonomous Systems. Proceedings der "First European Conference on Artificial Life" in Paris 1991, MIT Press 1992, 11-20

Paulin M. G. (1993): The Role of the Cerebellum in Motor Control and Perception. Brain Behavior and Evolution 41 (1993), 39-50

Peitgen, H.-O., Richter, P.H. (1986): The Beauty of Fractals. Springer, Berlin/Heidelberg

Pfeifer, R. (1995): Cognition. Erscheint in Robotics and Autonomous Systems

Piaget, J. (1947): La psychologie de l'intelligence. Librairie Armand Colin, Paris 1947. Deutsch: Psychologie der Intelligenz. Kindler, München 1974[6]

Piaget, J. (1968): Le structuralisme. Presses Universitaires de France, Paris 1968. (deutsch: Der Strukturalismus. Walter-Verlag, Olden/Freiburg 1982[2])

Popper, K.R., Eccles, J.C. (1977): The Self and Its Brain - An Argument for Interactionism. Springer Verlag, Heidelberg etc., 1977. Deutsch: Das Ich und sein Gehirn. Piper, München/Zürich 1982[2]

Port, R.F., Cummins, F., McAuley, J.D. (1994): Naive Time, Temporal Patterns and Human Audition. Research Report 118, Cognitive Science Program der Indiana University at Bloomington, Indiana, 1994. Erscheint auch in van Gelder & Port 1995

Powers, W.T. (1988): An Outline of Control Theory. In: Powers, W.T. (1989): Living Control Systems: Selected Papers of William T. Powers. The Control Systems Group Inc., 253-293

Preißl, H., Aertsen, A., Palm, G. (1990): Are Fractal Dimensions a Good Measure for Neural Activity? In: Eckmiller, R., Hartmann, G., Hauske, G. (Hrsg.) (1990): Parallel Processing in Neural Systems and Computers. North-Holland, Amsterdam 1990, 83-86

Prigogine, I. (1980): From Being to Becoming. Freeman, San Francisco. Deutsch: Vom Sein zum Werden. Piper, München 1979

Quillian, M.R. (1968): Semantic Memory. In: Minsky, M. (Hrsg.): Semantic Information Processing. MIT Press, Cambridge, Mass., 227-270

Rechenberg, I. (1973): Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. Friedrich Frommann Verlag, Stuttgart-Bad Cannstadt

Reignier, P. (1994): Fuzzy Logic Techniques for Mobile Robot Obstacle Avoidance. Robotics and Autonomous Systems 12 (1994), 143-153

Reiss, M., Taylor, J.G. (1991): Storing Temporal Sequences. Neural Networks 4(6), 773-781

Richards, D.D. (1988): Dynamic Concepts and Functionality: The Influence of Multiple Representations and Environmental Constraints on Categorization. Human Development 31, 11-19

Rickheit, G., Strohner, H. (1992): Towards a Cognitive Theory of Linguistic Coherence. Theoretical Linguistics 18 (2/3), 209-237

Riedel, U., Kühn, R., van Hemmen, J.L. (1988): Temporal Sequences and Chaos in Neural Nets. Physical Review A 38(2), 1105-1108

Rieger, B. (1985) (Hrsg.): Dynamik in der Bedeutungskonstitution. Buske, Hamburg

Robertson, S.S., Cohen, A.H., Mayer-Kress, G. (1993): Behavioral Chaos: Behind the Metaphor. In: Smith & Thelen 1993a, 119-150

Rojas, R. (1994): Approximationstheorie, Statistik und neuronale Netze. In: Duwe, I., Kurfeß, F., Paaß, G., Vogel, S. (Hrsg.) (1994): Konnektionismus und Neuronale Netze. Beiträge zur HeKoNN94. GMD-Studien 242, Gesellschaft für Mathematik und Datenverarbeitung (GMD), St. Augustin, 23-38

Sacks, E. (1990): A Dynamics System Perspective on Qualitative Simulation. Artificial Intelligence 42 (1990), 349-362

Sacks, E. (1991): Automatic Analysis of One-parameter Planar Ordinary Differential Equations by Intelligent Numerical Simulation. Artificial Intelligence 48 (1991), 27-56

Sandewall, E. (1994): Features and Fluents: The Representation of Knowledge about Dynamical Systems, Vol. 1. Clarendon Press, London

Schank, R.C. (1982): Dynamic Memory. Cambridge University Press, Cambridge 1982

Scheerer, E. (1990): Konnektionismus und Symbolverarbeitung: Einige Traditionslinien in der deutsche Psychologie. Report Nr. 59 der Forschungsgruppe "Mind and Brain", Zentrum für interdisziplinäre Forschung (ZIF), Universität Bielefeld

Scheier, Ch., Pfeifer, R. (1995): Classification as Sensory-Motor Coordination: A Case Study on Autonomous Agents. Erscheint in den Proceedings der ECAL-95

Schmidt, R.F. (Hrsg.) (1980^4): Grundriß der Sinnesphysiologie. Springer Verlag, Berlin/Heidelberg/New York

Schmidt, S.J. (Hrsg.) (1987a): Der Diskurs des Radikalen Konstruktivismus. Suhrkamp, Frankfurt (stw 636)

Schmidt, S.J. (Hrsg.) (1987b): Der Radikale Konstruktivismus: Ein neues Paradigma im interdisziplinären Diskurs. In Schmidt, S.J. (1987a), 11-88

Schöner, G., Dose, M. (1992): A Dynamical Systems Approach to Task-Level System Integration Used to Plan and Control Autonomous Vehicle Motion. Robotics and Autonomous Systems 10, 253-267

Schöner, G., Haken, H., Kelso, J.A.S. (1986): A Stochastic Theory of Phase Transitions in Human Hand Movement. Biological Cybernetics 53, 247-257

de Schutter, E., Bower, J.M. (1994): An Active Membrane Model of the Cerebellar Purkinje Cell. Journal of Neurophysiology 71 (1), 375-400 (Part 1), 401-419 (Part 2)

Schwefel, H.-P. (1977): Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie. Birkhäuser, Basel/Stuttgart 1977

Schweitzer, G., Wen, J. (1994): Where Neural Nets Make Sense in Robotics. In: Proceedings der Konferenz "Prerational Intelligence in Robotics: from Sensorimotor Intelligence to Collective Behavior", Report Nr. 10 der Forschungsgruppe "Prerational Intelligence", Zentrum für interdisziplinäre Forschung (ZIF), Universität Bielefeld, 109-135

Seidenberg, M.S., Tanenhaus, M.K., Leiman, J.M., Bienkowski, M. (1982): Automatic Access of the Meaning of Ambiguous Words in Contexts: Some Limitations of Knowledge-Based Processing. Cognitive Psychology 14, 489-537

Shimohara, K. (1994): Evolutionary Systems for Brain Communications - Towards an Artificial Brain. In Brooks, R., Maes, P. (Hrsg.) (1994): Artificial Life IV. Proceedings des Fourth Int. WS on the Synthesis and Simulation of Living Systems. Bradford/MIT Press, 3-7

Sims, K. (1994): Evolving 3D Morphology and Behavior by Competition. In Brooks, R.A., Maes, P. (1994): Artificial Life IV. Proc. des Fourth International Workshop on the Simulation and Synthesis of Living Systems. Bradford Books/MIT Press, 28-39

Sloman, A. (1975): Afterthoughts on Analogical Representations. Proceedings "Theoretical Issues in Natural Language Processing", Cambridge, Mass., 1975, 164-168. Nachdruck in Brachman, R.J., Levesque, H.J. (Hrsg.) (1985): Readings in Knowledge Representation. Morgan Kaufman, Los Altos, 431-439

Smith, E.E., Medin, D.L. (1981): Categories and Concepts. Harvard University Press, Cambridge, Mass.

Smith, L.B., Thelen, E. (Hrsg.) (1993a): A Dynamic Systems Approach to Development: Applications. Bradford/MIT Press, Cambridge, Mass.

Smith, L.B., Thelen, E. (1993b): From the Dynamics of Motor Skill to the Dynamics of Devlopment. In: Smith & Thelen 1993a, 1-11

Smith, L.B., Thelen, E. (1993c): Can Dynamic Systems Theory be Usefully Applied in Areas Other than Motor Development? In: Smith & Thelen 1993a, 151-170

Smithers, T. (1994): What the Dynamics of Adaptive Behavior and Cognition Might Look Like in an Agent-Environment Interaction System. In: Proceedings des Workshops "On the Role of Dynamics and Represenation in Adaptive Behavior and Cognition", Universidad del Pais Vasco, San Sebastian 1994, 135-153

Smolensky, P.(1986): Information Processing in Dynamical Systems: Foundations of Harmony Theory. In: Rumelhart, D.E., McClelland, J.L. (Hrsg.): Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1, MIT Press, Cambridge, Mass. 1986, 194-281

Stadler, M., Kruse, P. (1986): Gestalttheorie und Theorie der Selbstorganisation. Gestalt Theory 8, 75-98

Stadler, M., Kruse, P. (1990): The Self-Organization Perspective in Cognition Research: Historical Remarks and New Experimental Approaches. In: Haken, H., Stadler, M. (Hrsg.) (1990): Synergetics of Cognition. Springer Series in Synergetics 45, Springer Verlag, 32-52

Steels, L. (1993a): Building Agents out of Autonomous Behavior Systems. In: Steels, L., Brooks, R. (Hrsg.) (1993): The "Artificial Life" Route to "Artificial Intelligence": Building Situated Embodied Agents. Lawrence Erlbaum, New Haven 1993

Steels, L. (1993b): A Mathematical Framework for Autonomous Robots. Proc. der EWAIC 93, Moskau 1993, 333-334

Steels, L. (1994a): The Artificial Life Roots of Artificial Intelligence. Erscheint im Artificial Life Journal 1(1)

Steels, L. (1994b): Mathematical Analysis of Behavior Systems. In: Gaussier, P., Nicoud, J.-D. (Hrsg.): Proceedings of the "From Perception to Action Conference", Lausanne, Sept. 1994, IEEE Computer Society Press, Los Alamitos, CA, 88-95

Steels, L. (1994c): Emergent Functionality in Robotic Agents through On-Line Evolution. In Brooks, R., Maes, P. (Hrsg.) (1994): Artificial Life IV. Proceedings des Fourth Int. WS on the Synthesis and Simulation of Living Systems. Bradford/MIT Press, 8-14

Steels, L. (1994d): A Case Study in the Behavior-Oriented Design of Autonomous Agents. In: Cliff, D. et al. (Hrsg.) (1994): From Animals to Animats III. Proceedings der dritten internationalen Konferenz "On the Simulation of Adaptive Behavior", Bradford/MIT Press, 445-452

Steinkühler, U., Cruse, H. (1994): A Holistic Model for the Control of an Arm with Redundant Degrees of Freedom. In: Proceedings der Konferenz "Prerational Intelligence in Robotics: from Sensorimotor Intelligence to Collective Behavior", Report Nr. 10 der Forschungsgruppe "Prerational Intelligence", Zentrum für interdisziplinäre Forschung (ZIF), Universität Bielefeld, 149-163

Stengel, R.F. (1986): Stochastic Optimal Control. Wiley, New York

Stoecker, M., Eckhorn, R. (1994): Synthesizing Complex Perceptions II: Models of Visual Feature Associations Based on Neural Synchronization. Proceedings der Konferenz "Emergence of Prerational Intelligence in Biology: From Sensorimotor Intelligence to Collective Behavior", Part 1, Report Nr. 7 der Forschungsgruppe "Prerational Intelligence", Zentrum für interdisziplinäre Forschung (ZIF), Universität Bielefeld, 101-114

Suchman, L.A. (1987): Plans and Situated Actions: The Problem of Human-Machine Communication. Cambridge University Press, Cambridge, Mass., 1987

Tani, J. (1995): Embedding a Grammatical Description in Deterministic Chaos: An Experiment in Recurrent Neural Learning. Erscheint in Biological Cybernetics

Thomas, R. (Hrsg.) (1979): Kinetic Logic: A Boolean Approach to the Analysis of Complex Regulatory Systems. Springer Verlag, Berlin/Heidelberg/New York 1979 (Lecture Notes in Biomathematics 29)

Thomas, R., Nicolis, G., Richelle, J., Van Ham, P. (1979): General discussion on the simplifying assumptions in methods using logical, stochastic or differential equations; the ranges of applicatbility and the complementarity of the approaches. In Thomas, R., (1979), 345-352

Tino, P., Horne, B.G., Giles, C.L. (1995): Finite State Machines and Recurrent Neural Networks - Automata and Dynamical Systems Approach. Technical Report, UMIACS-TR-95-1 und CS-TR-3396, Institute for Advanced Computer Studies, University of Maryland

Toffoli, T. (1981): Physics and Computation. Int. Journal of Theoretical Physics 21 (3/4), 1982, 165-175

Townsend, J.T., Busemeyer, J.R. (1989): Approach-Avoidance: Return to Dynamic Decision Behavior. In: Izawa, C. (Hrsg.) (1989): Current Issues in Cognitive Processes. Lawrence Erlbaum, 107-133

Treisman, A. (1986): Properties, Parts, and Objects. In: Boff, K. R., Kaufman, L., und Thomas, J.P. (Hrsg.): Handbook of Perception and Human Performance, vol. 2. John Wiley and Sons, New York 1986, Kapitel 35

Triggs, B. (1994): Model-based Sonar Localisation for Mobile Robots. Robotics & Autonomous Systems 12, 173-186

Tucker, M., Hirsh-Pasek, K. (1993): Systems and Language: Implications for Acquisition. In Smith & Thelen 1993a, 359-384

Ulrich, H., Probst, G.J.B. (1984) (Hrsg.): Self-Organization and Management of Social Systems. Springer Verlag, Berlin/Heidelberg/New York/Tokyo

Vera A.H., Simon, H.A. (1993a): Situated Action: A Symbolic Interpretation. Cognitive Science 17 No. 1, 1993, 7-48

Vera A.H., Simon, H.A. (1993b): Situated Action: Reply to William Clancey. Cognitive Science 17 No. 1, 1993, 117-135

Verschure, P.F.M.J. (1993): The Cognitive Development of an Autonomous Behaving Artifact: The Self-Organization of Categorization, Sequencing, and Chunking. Proceedings der Konferenz "Prerational Intelligence - Phenomenology of Complexity Emerging in Systems of Agents Using Simple Rules", Report Nr. 2 der Forschungsgruppe "Prerational Intelligence", Zentrum für interdisziplinäre Forschung (ZIF), Universität Bielefeld, 95-120

Wagner, G. (1986): Evolution der Evolutionsfähigkeit. In Dress et al. 1986, 121-147

Waltz, D. L., Pollack, J.B. (1985): Massively Parallel Parsing: A Strongly Interactive Model of Natural Language Interpretation. Cognitive Sience 9, 51-74.

Weidemann, H.-J., Pfeiffer, F. (1994): The Control System of the Six-Legged TUM Walking Robot. In: Proceedings der Konferenz "Prerational Intelligence in Robotics: from Sensorimotor Intelligence to Collective Behavior", Report Nr. 10 der Forschungsgruppe "Prerational Intelligence", Zentrum für interdisziplinäre Forschung (ZIF), Universität Bielefeld, 165-168

von Weizsächer, E.U. (1987): Brückenkonzepte zwischen Natur- und Geisteswissenschaften: Selbst-organisation, Offene Systeme und Evolution. Sozial-ökologische Arbeitspapiere 17, Forschungsgruppe Soziale Ökologie, Frankfurt/Main 1987, 1-27

Wellek, A. (1972): Ganzheit, Gestalt und Nichtgestalt. Wandel und Grenzen des Gestaltbegriffs und der Gestaltkriterien. In: Weinhandl, F. (Hrsg.): Gestalthaftes Sehen. Wissenschaftliche Buchgesellschaft, Darmstadt 1978, 384-397

Wiener, N. (1961²): Cybernetics. MIT Press, Cambridge, Mass.

Wildgen, W. (1985): Archetypensemantik. Grundlagen für eine dynamische Semantik auf der Basis der Katastrophentheorie. Gunter Narr Verlag, Tübingen

Wilks, Y. (1975): A Preferential, Pattern-Seeking Semantics for Natural Language Inference. Artificial Intelligence 6, 1975, 53-74

Willke, Helmut (1987): Systemtheorie, 2. Auflage. Gustav Fischer Verlag, Stuttgart 1987 (UTB 1161)

Winograd, T., Flores, F. (1986): Understanding Computers and Cognition. Ablex 1986. Deutsch: Erkenntnis Maschinen Verstehen. Rotbuch Verlag, Berlin 1989

Wolff, P.H. (1993): Behavioral and Emotional States in Infancy: A Dynamic Perspective. In Smith & Thelen 1993a, 189-208

Wolfram, S. (1983): Statistical Mechanics of Cellular Automata. Rev. of Mod. Physics 55 (3), 1983, 601-644

Wolfram, S. (1984): Undecidability and Intractability in Theoretical Physics. Physical Review Letters 54 (8), 1985, 735-738

Wolfram, S. (1986) (Hrsg.): Theory and Applications of Cellular Automata. World Scientific, Singapore 1986

Wunsch, Gerhard (1985): Geschichte der Systemtheorie. Oldenbourg, München 1985

Yamauchi, B., Beer, R. (1994): Integrating Reactive, Sequential, and Learning Behavior Using Dynamical Neural Networks. In: Cliff, D. et al. (Hrsg.) (1994): From Animals to Animats III. Proc. der dritten internationalen Konferenz "On the Simulation of Adaptive Behavior", Bradford/MIT Press, 382-391

Yao, Y., Freeman, W.J. (1990): A Model of Biological Pattern Recognition with Spatially Chaotic Dynamics. Neural Networks 3, No. 2, 153-170

Yao, Y., Freeman, W.J., Burke, B., Yang, Q. (1991): Pattern Recognition by a Distributed Neural Network: An Industrial Application. Neural Networks 4 No.1 (1991), 103-121

Zadeh, L.A. (1969): The Concept of System, Aggregate, and State in System Theory. In Zadeh & Polak 1969, 3-42

Zadeh, L.A., Polak, E. (Hrsg.) (1969): System Theory. McGraw-Hill, New York 1969 (Inter-University Electronics Series Vol. 8)

Zalama, E., Gaudiano, P., Coronado, J.L. (1995): A Real-Time, Unsupervised Neural Network for the Low-Level Control of a Mobile Robot in a Nonstationary Environment. Neural Networks 8 (1), 103-123

Zwölfer, H. (1986): Insektenkomplexe an Disteln - ein Modell für die Selbstorganisation ökologischer Kleinsysteme. In: Dress et al. 1986, 181-218