Repräsentation ohne Repräsentation - Überlegungen zu einer Neurodynamik modularer kognitiver Systeme*

Frank Pasemann

1 Einführung

Das biologische Gehirn als eines der komplexesten und faszinierendsten Systeme, dem die Wissenschaft bislang ihre Aufmerksamkeit zugewandt hat, besteht aus Myriaden von Zellen, die selbst wieder sehr komplexe Gebilde sind. Offenbar ermöglicht deren wohlgeordnetes Zusammenwirken, daß höhere Lebewesen fähig sind, ihre Aktionen flexibel auf die eigenen Bedürfnisse und die Bedingungen ihrer Umwelt abzustimmen. Die beeindruckenden Fähigkeiten von Gehirnen werfen ein breites Spektrum an Fragestellungen auf; entsprechend nähert sich Hirnforschung ihrem Gegenstand auf sehr unterschiedlichen Ebenen. Von der Molekularbiologie der Zelle bis zur Verhaltensebene von Lebewesen reichen Untersuchungen, an denen Disziplinen wie z.B. die Psychologie, Neurobiologie, Neurophysiologie und die Kognitionswissenschaften wesentlich beteiligt sind und zu denen Neuroinformatik, Physik und Mathematik theoretische Beiträge liefern. Bei diesem multidisziplinären Forschungsbemühen ist es nicht verwunderlich, daß die bislang verwendeten Begriffe oft unscharf, in steter Neudefinition und Hinterfragung begriffen sind: Wahrnehmung, Information, Selbstorganisation, Selbstreferenz, Emergenz, Komplexität, Bewußtsein, Gedächtnis, Lernen, Aufmerksamkeit und nicht zuletzt der Begriff der inneren Repräsentation sind Beispiele dafür. Hinzukommt eine starke Fraktionierung der Erkenntnisse in Teilergebnisse verschiedenster Disziplinen, die zusammen gesehen und gedacht werden müssen. Neue Technologien wie die Patch-clamp Technik auf der Zellebene, Multielektrodenableitungen, EEG (electroencephalogram) und bildgebende Verfahren wie MEG (magnetoencephalogram), PET (positron emission tomography) und funktionales MRI (magnetic resonance imaging) (vgl. z.B. Thatcher et.al. 1994; Mazoyer et.al. 1995) erlauben die Präzisierung, Vertiefung und Diversifizierung der Fragestellungen auf verschiedenen Systemebenen und tragen zu dem momentanen Bild

* in: G. Rusch, S. J. Schmidt und O. Breidbach, *Interne Repräsentationen - Neue Konzepte der Hirnforschung*, Frankfurt: Suhrkamp (stw 1277), 1996.

unübersichtlicher Einzelergebnisse bei. Die Fragen: "Was tun Gehirne, um einem Lebewesen das Überleben in einer sich permanent verändernden Umwelt zu ermöglichen, wann und wie tun sie es?" sind erst in Teilen zu beantworten.

Als eine für die Analyse und das Verständnis von kognitiven Hirnprozessen wesentliche Grundfrage wird die nach der "inneren Repräsentation" der externen Welt in Gehirnen angesehen. Vorstellungen von der Form dieser Repräsentationen sind eng verknüpft mit detaillierten Kenntnissen über die Struktur des Gehirns bzw. die Teilfunktionen bestimmter Gehirnareale, sowie mit der Frage nach dem Gedächtnis, seiner Realisierung und der Art seiner Verwendung im Zusammenhang mit einem "sinnvollen" Verhalten von Lebewesen. Der Kreis schließt sich mit der Frage nach den Lernprozessen, d.h. nach der Beschaffenheit von strukturellen und funktionalen Prozessen, in denen sich innere Repräsentationen und Gedächtnis ausprägen. Auch auf diese Fragen gibt es zur Zeit weder überzeugende Antworten noch geschlossene theoretische Konzepte.

Eines der hervorstechenden funktionalen Merkmale von Gehirnen, wenn wir zunächst die menschliche Erfahrung zugrunde legen, ist ihre Fähigkeit, nicht nur passiv auf Signale der externen Welt zu reagieren, sondern darüber hinaus Situationen vorherzusehen, Intentionen zu verfolgen, Pläne für ein Verhalten zu entwickeln, dieses zu kontrollieren, zu evaluieren und so zu verändern, daß es mit den Intentionen und Plänen korrespondiert. Es sind offenbar diese *kognitiven Fähigkeiten* von Gehirnen, die komplexen Organismen dazu verhelfen, ihre Aktionen und Bedürfnisse an die stetig wechselnden Bedingungen einer vorgefundenen Umwelt so anzupassen, daß sie überleben können. Ein Lebewesen muß fähig sein, gewisse Veränderungen in seiner Umgebung korrekt vorherzusagen, um verhaltensrelevante Entscheidungen zu fällen. Verbreitet wird heute der Standpunkt eines externen Beobachters eingenommen, der einem solchen Lebewesen eine angemessene "innere Repräsentation" seiner Umwelt zuschreibt, auf deren Grundlage es ein prädiktives "inneres Weltmodell" entwickelt, "Handlungspläne" erstellt und "Ziele" verfolgt, die es ihm ermöglichen, unter den gegebenen Randbedingungen seine Existenz zu sichern und sich zu reproduzieren. Kognitive Prozesse in Gehirnen "operieren" nach dieser Vorstellung auf oder mit inneren Repräsentationen von Objekten und Situationen der externen Welt. Diese Anschauung entspricht weitgehend jener Methodik prädiktiver "Weltmodellierung", wie sie in unserer bisherigen wissenschaftlichen Praxis erfolgreich angewandt wird. Dennoch ist die Frage zu stellen, ob der Begriff der inneren Repräsentation nicht nur ein meta-sprachliches Beschreibungselement ist, dem auf der Ebene biologischer Hirnprozesse kein materielles Korrelat entspricht. Wir werden dieser Frage in dem vorliegenden Artikel nachgehen.

Im folgenden verstehen wir unter einem *kognitiven System* ein cerebrales Teilsystem, das in der Lage ist, komplexe lokale Operationen auszuführen, deren Zusammenwirken in einem kognitiven Prozeß jene Fähigkeiten hervorbringt, die wir oben als kognitive Fähigkeiten charakterisiert haben. Im Rahmen einer in den letzten 30 Jahren stark von

der Künstlichen Intelligenz (KI) beeinflußten Sehweise werden solche Systeme bisweilen auch als "höhere" informationsverarbeitende Systeme beschrieben, die an ihren Eingängen Information aus der Umwelt empfangen und an ihren Ausgängen Information zur Kontrolle der Motorik bereitstellen. Verknüpft mit dieser Vorstellung sind innere Repräsentationen der externen Welt in der Form von Symbolen, einem Gedächtnis als Informationsspeicher und Kognition als einem symbolverabeitenden Prozeß. Auf diesem Ansatz beruht die sogenannte Computer-Metapher des Gehirns. Bereits der in den 80er Jahren aufkommende Konnektionismus (Rumelhart und McClelland 1986), der sich stärker an dem neuronalen Substrat biologischer Gehirne orientiert, betont die Entstehung von kognitiven Fähigkeiten in einem selbstorganisatorischen Lernprozeß. Kognition als informationsverarbeitender Prozeß beruht hier jedoch auf sogenannten *sub-symbolischen* Repräsentationen, die in Form der gewichteten synaptischen Verbindungen in einem trainierten neuronalen Netz vorliegen.

Betrachtungsweisen, die Gehirne als reine informationsverarbeitende Systeme verstehen, werden heute zunehmend durch einen Ansatz abgelöst bzw. ergänzt, der auf Vorstellungen der mathematischen Theorie dynamischer Systeme (vgl. z.B. Wiggins 1990) aufbaut. Dies beruht weitgehend auf der Einsicht, daß der in der Shannonschen Informationstheorie begründete Informationsbegriff zur Analyse und Beschreibung von Hirnprozessen inadäquat ist, vornehmlich, weil die Informationstheorie sich dem Problem der "Bedeutung" von Signalen kaum zuwendet (vgl. z.B. Roth (1994), S. 92f). Bedeutung ist in diesem Kontext eher ein biologisches Problem. Denn in bezug auf die kognitiven Fähigkeiten von Gehirnen ist die eigentlich interessante Frage nicht: "Wie verarbeiten Gehirne Information?", sondern: "Wie erzeugen Gehirne Information, oder besser, Bedeutung?" Dies meint, daß äußere Signale, innere Prozesse und insbesondere innere Repräsentationen Bedeutung nur dann und insofern erlangen, als sie für das Verhalten eines Lebewesens relevant sind, also letztlich die Überlebensfähigkeit eines Wesens in seiner Umwelt ermöglichen (Roth 1992). Der dynamische Zugang zu dieser Problematik, der die Fähigkeit zur Selbstorganisation in einem steten Wechselwirkungsprozeß mit der Umwelt als entscheidende und charakteristische Eigenschaft kognitiver Systeme zur Voraussetzung hat, ist erst in Umrissen erkennbar. Unter synergetischen Gesichtspunkten wird er z.B. von Kelso (1995) und Haken (1995) skizziert. Er wird aber auch zunehmend in den Kognitionswissenschaften diskutiert (Port und van Gelder 1995), sowie in Forschungsgebieten mit vergleichbaren Fragestellungen, wie z.B. dem der "Autonomen Agenten" (Steels und Brooks 1995) und der "Komplexen Adaptiven Systeme" (Holland 1995).

In bezug auf die folgenden Betrachtungen sei kurz auf die Organisationsstruktur von biologischen Hirnen eingegangen (vgl. z.B. Roth 1994). Die spezifischen Fähigkeiten eines Gehirns beruhen im wesentlichen auf den Wechselwirkungen einer sehr großen Zahl von Nervenzellen, den Neuronen, die vermöge sogenannter synaptischer Verbindungen erfolgen. Dabei ist zu berücksichtigen, daß unter funktionalen Gesichtspunkten verschiedene Areale in Gehirnen zu unterscheiden sind. So sind

Bereiche von Neuronen, die wesentlich zur Verarbeitung von visuellen, akustischen, olfaktorischen oder taktilen Signalen beitragen, an verschiedenen Orten im Gehirn lokalisierbar, ebenso wie Areale, die bei der Erzeugung der verschiedenen motorischen Signale mitwirken. Folgt man dem Signalfluß von sensorischen zu motorischen Bereichen, so scheinen die Areale hierarchisch geordnet zu sein, z.B. von der Retina über den Corpus geniculatem laterale zum primären visuellen Cortex und weiter zu den höheren visuellen Arealen. Für unsere spätere Argumentation ist aber entscheidend, daß es eine Vielzahl von Rückkopplungen oder auch "Rückprojektionen" gibt, die zum Teil ebenso stark ausgeprägt sind wie die vorwärts gerichteten Verbindungen. Fast alle Bereiche des Gehirns wirken durch die Existenz solcher geschlossenen Signalschleifen auf sich selbst zurück. Gleiches gilt natürlich auch für die einzelnen Neuronen, die innerhalb eines Areals oder in räumlich getrennten Bereichen des Gehirns an der Signalverarbeitung beteiligt sind. Außerdem gibt es Verbindungen, die auf die folgenden Neuronen exzitatorisch (erregend) wirken, andere die inhibitorisch (hemmend) sind. Darüber hinaus ist sowohl die Divergenz von Verbindungen zu beobachten, d.h. ein Neuron (Areal) sendet seine Signale an viele andere, als auch deren Konvergenz, d.h. ein Neuron (Areal) erhält Signale von vielen anderen Neuronen (Arealen). Die starke Divergenz und Konvergenz von interneuronalen Verbindungen hat letztlich die Vorstellung von einer hochgradig parallelen Signalverarbeitung in Gehirnen begründet.

Es ist insbesondere die erwähnte, ausgeprägt rekursive Struktur der neuronalen "Verschaltung" von Gehirnen, die zusammen mit den nichtlinearen Eigenschaften der einzelnen Neuronen unter dynamischen Gesichtspunkten ein sehr komplexes Verhalten neuronaler Systeme erwarten läßt. Tatsächlich zeugen neuere Ergebnisse der Hirnforschung, insbesondere die "Synchronisation" oszillierender Neuronen als Antwort auf äußere Reize (vgl. Artikel in Krüger 1991; Buzsáki et.al. 1994; Pantev et.al. 1995) sowie das Auffinden chaotischer Dynamik in bestimmten Hirnarealen (vgl. z.B. Elbert et.al. 1994, und Artikel in Duke und Pritschard 1991), von der Existenz komplexer dynamischer Prozesse in Gehirnen. Die bei unseren Überlegungen zu berücksichtigende Hirndynamik betrifft sowohl die Veränderung neuronaler Aktivität als auch die Veränderung der Synapsenstärken, d.h. der Stärke der interneuronalen Verbindungen. Wir verstehen sie hier letztlich als Ausdruck eines Selbstorganisationsprozesses, der durch Stabilität und Instabilität, durch Kooperation und Konkurrenz seiner Teilprozesse zu charakterisieren ist. Grundlage für unsere Überlegungen ist daher die Beschreibung kognitiver Systeme als spezifische nichtlineare dynamische neuronale Systeme.

Für unsere Argumentation bezüglich der inneren Repräsentation äußerer Welt wird wichtig sein, kognitive Systeme nicht als isoliert zu betrachten, sondern als Systeme, die in eine sogenannte *sensomotorische Schleife* eingebunden sind. Diese Einbindung setzt entscheidende Randbedingungen sowohl für die Entstehung der lokalen sowie der globalen neuronalen Organisationsstrukturen als auch für die funktionalen Teilprozesse und deren Interaktion in Gehirnen. Wie wir ausführen werden hängen

Organisationsstruktur und Teilfunktionen kognitiver Systeme auch von der physischen Beschaffenheit des "Wahrnehmungsapparates" (Sensorik) und des "Bewegungsapparates" (Motorik) des Lebewesens ab.

Bevor wir darangehen zu untersuchen, ob und in welcher Form innere Repräsentationen der externen Welt in Gehirnen für das Verhalten von Lebewesen konstitutiv sind, werden wir zunächst einige Eigenschaften kognitiver Systeme umreißen, um dann wichtige Grundbegriffe aus der Theorie dynamischer Systeme für die folgenden Diskussionen bereitzustellen. Die Möglichkeiten einer komplexen biologischen Neurodynamik werden wir an Beispielen aus dem Bereich kleiner *künstlicher* neuronaler Einheiten, den Neuromodulen, verdeutlichen. Die erwähnte Lokalisierbarkeit einzelner, bestimmten Funktionen zugeordneter Hirnbereiche legt den modularen Aufbau von Gehirnen bereits nahe. Unser Rekurs auf Module als Grundelemente eines kognitiven Systems ist jedoch methodisch bedingt und erlaubt insbesondere, Multifunktionalität bzw. funktionale Flexibilität schon auf der untersten Beschreibungsebene des Systems einzuführen: ein Vorgehen, das der Funktionsweise neuronaler Systeme durchaus adäquat ist. Unter dem dann erläuterten Begriff des *autotropen Systems* fassen wir unsere Vorstellungen von der Entfaltung kognitiver Fähigkeiten in einem Selbstorganisationsprozeß neuromodularer Strukturen zusammen. Der Verweis auf solche "auf sich selbst Einfluß nehmenden" Systeme ist vielen Beschreibungsansätzen gemein, die die Selbstorganisation als Grundlage des Verhaltens komplexer Systeme betrachten (z.B. Maturana und Varela 1987; Haken 1995; Kelso 1995). Wir möchten in unserer Argumentation jedoch schon fixierte Begriffe wie "Synergetik" oder "Autopoiesis" vermeiden, um die spezifischen Eigenschaften *modularer* neuronaler Systeme deutlicher herausarbeiten zu können. Auf diesem Hintergrund wird deutlich werden, daß das "klassische" Konzept innerer Repräsentation der externen Welt nur schwerlich aufrechtzuerhalten ist. Ausgehend von der Annahme, daß der Terminus "innere Repräsentation" zur Charakterisierung der Grundlage handlungsrelevanter Entscheidungen dennoch sinnvoll ist, werden wir versuchen, ihn inhaltlich neu zu fassen.

2 Kognitive Systeme

Da wir kognitiven Systemen zuschreiben wollen, innere Repräsentationen zu besitzen, interessieren uns zunächst jene Prozesse, die an der Hervorbringung kognitiver Fähigkeiten von Gehirnen besonderen Anteil haben. Dazu denken wir uns das Gehirn als System grob in drei funktional unterschiedene Bereiche partitioniert, in das sensorische, motorische und das kognitive Teilsystem. Die genaue Bestimmung dieser Trisektion ist für die folgenden Betrachtungen nicht relevant, aber es ist davon auszugehen, daß der Neocortex, im Zusammenspiel mit vielen anderen Bereichen, einen wesentlichen Anteil an der Realisierung kognitiver Fähigkeiten hat (vgl. z.B. Roth (1994), Kap. 9). Das sensorische Teilsystem umfaßt die Rezeptoren der verschiedenen

Sinnesmodalitäten und Areale, welche die sensorseitigen Signale geeignet aufbereiten und das Ergebnis u.a. an das kognitive System weiterleiten. Das motorische System setzt die u.a. vom kognitiven System empfangenen Signale in geeignete Befehle an die Effektoren um. Wichtig sind aber ebensosehr die "rückwärts" gerichteten Verbindungen, d.h. das kognitive System sendet im allgemeinen Signale auch ans sensorische System, vermutlich im Zusammenhang mit Selektions- bzw. Aufmerksamkeitsmechanismen, und empfängt Signale vom motorischen System über dessen momentanen Zustand. Die funktionale Trisektion des Gehirns ist hier also nicht als signalführende "Einbahnstraße" von den Sensoren über ein Zentralsystem zu den Effektoren zu verstehen. Wie oben erwähnt soll das kognitive Teilsystem Eigenschaften zeigen, die wir in der Regel "mentalen" Aktivitäten in (nicht notwendig menschlichen) Gehirnen zuschreiben. Ein kognitives System sollte etwa in der Lage sein, wesentliche Aspekte seiner Umwelt mittels der Sinnesorgane wahrzunehmen, zu generalisieren, zu kategorisieren, von ihnen zu abstrahieren und Relationen zwischen den Aspekten zu erkennen. Es muß fähig sein, verschiedene Arten von Gedächtnis (z.B. Kurz- und Langzeitgedächtnisse) auszubilden, um gewisse Veränderungen in der Umwelt zu konstatieren und vorherzusagen, und es muß Kriterien entwickeln, um verhaltensrelevante Entscheidungen fällen zu können. Ferner muß es über das motorische System die Möglichkeit haben, sich in seiner Umwelt zu bewegen und auf sie einzuwirken.

Aus dieser generellen Trisektion des Gehirns folgt sofort, daß ein kognitives System nicht als isoliertes System betrachtet werden kann. Seine Entstehung und Existenzberechtigung beruht auf der Existenz und der physischen Konstitution der peripheren Systeme. Wir gehen davon aus, daß sowohl die Struktur und die Funktionen kognitiver Systeme als auch die Prinzipien seiner Selbstorganisation ohne Berücksichtigung dieses Sachverhalts nicht zu verstehen sind. Die Eigenschaft, daß Kognition nur in bezug auf einen gegeben Körper, also in bezug auf eine gegebene Sensorik und Motorik zu verstehen ist, wird auch als "embodyment" bzw. "Verleiblichung" (Varela, 1994) bezeichnet. Sie setzt eine erste, *innere Randbedingung* an die Entfaltung spezifischer neuronaler Strukturen und deren Funktionen.

Eine weitere fundamentale Eigenschaft kognitiver Systeme besteht in ihrer Fähigkeit, sich an die Anforderungen einer sich permanent verändernden Umwelt zu adaptieren und aus der Erfahrung des eigenen Agierens zu lernen. Dies geschieht in der sensomotorischen Schleife letztlich mit dem Ziel in der gegebenen Umwelt möglichst gut und lange zu überleben. In dieser Schleife empfängt das Lebewesen Signale aus seiner Umwelt vermöge der Sensorik und agiert in oder auf die Umwelt mit Hilfe seiner Motorik. Dieses Eingebundensein in eine sensomotorische Schleife setzt eine zweite, *äußere Randbedingung* für die Selbstorganisation neuronaler Systeme und deren kognitive Prozesse. Struktur wie Funktion eines kognitiven Systems hängen also von der "ökologischen Nische" ab in der das Gesamtsystem zu überleben hat. Diese Eigenschaft bezeichnet man gelegentlich als "situatedness" oder "Situiertheit" (Varela, 1994).

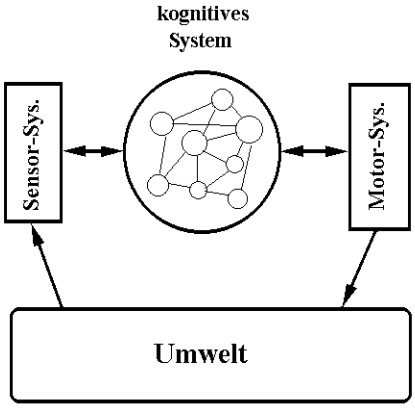


Abb. 1: Ein kognitives System in der sensomotorischen Schleife.

Selbstorganisation und Selbstreferenz sind als grundlegende Organisationsprinzipien eines kognitiven Systems anzusehen. Unter Selbstorganisation verstehen wir hier, daß das kognitive System, ausgehend von einem phylogenetisch vorgegebenen Repertoire an Verhaltensmöglichkeiten, aus eigener Erfahrung das seiner Umwelt angemessene Verhalten (im Sinne der Selbsterhaltung) erlernen muß. Der mathematisch-physikalische Begriff der Selbstorganisation bezieht sich auf das Verhalten komplexer Systeme, d.h. er setzt sowohl eine kritische (große) Zahl von Systemelementen als auch eine Vielzahl von "Verbindungen" voraus, vermöge derer eine nichtlineare Wechselwirkung zwischen ihnen erfolgt. Das Ergebnis solch lokaler Interaktionen bringt neue Eigenschaften hervor, die weder an den einzelnen Elementen beobachtet noch als pure Addition von Elementeigenschaften verstanden werden können. Dies ist wesentlich auf die Nichtlinearität der Wechselwirkungen zurückzuführen. Entscheidend für einen Selbstorganisationsprozeß ist die Balance von kooperierenden und konkurrierenden Systemteilen, von Stabilität und Instabilität der Teilprozesse. Der Begriff der Emergenz verweist auf die dabei entstehenden globalen raum-zeitlichen Ordnungsstrukturen, die für einen Selbstorganisationsprozeß typisch sind. *Selbstreferenz* verweist auf die Tatsache, daß ein Selbstorganisationsprozeß nicht monokausal verläuft, d.h. daß die Entstehung globaler Eigenschaften des Systems auf der lokalen Interaktion von Elementen beruht, die lokalen Interaktionen jedoch wiederum durch diese globalen Eigenschaften bestimmt sind. Diese Eigenschaft wird oft auch als "zyklische Kausalität" oder als "operationale Geschlossenheit" bezeichnet. In diesem Sinne ist auch die sensomotorische Interaktion des Lebewesens mit seiner Umwelt selbstreferenziell, weil die Veränderung der Sinneswahrnehmung u.a. durch dessen Bewegung, und seine Bewegung durch dessen Sinneswahrnehmung bestimmt wird.

Die Selbstorganisation kognitiver Systeme wird auf zwei verschiedene Weisen beeinflußt: Einmal durch die über die Sensorik aufgenommenen Signale aus der äußeren Umwelt und die durch die Motorik erzeugten Signale, zum anderen durch die eigenen,

intern ablaufenden Prozesse, die letztlich das Verhalten bestimmen. Bedingt durch die beiden genannten Einflüsse sind sie in der Lage sowohl die Struktur als auch die Funktion entsprechender Untereinheiten auf gezielte Weise zu verändern. Unter Strukturänderung in einer Untereinheit verstehen wir hier das Hinzufügen oder Beseitigen von Neuronen bzw. von synaptischen Verbindungen zwischen Neuronen sowie die Variation entsprechender Synapsenstärken. Der Selbstorganisationsprozeß verläuft außerdem auf drei kausal aufeinander aufbauenden unterschiedlichen Zeitebenen. Auf der unteren Ebene wird sich die Verschaltung, d.h. die anatomische Architektur herausbilden: Module und Aggregate von Modulen werden geformt. Ein Prozeß, den wir als Adaptation bezeichnen. Auf einer zweiten Zeitskala, die wir den spezifischen Lernprozessen zuordnen, werden die Synapsenstärken der Verbindungen variiert. Die dritte Zeitskala betrifft die kognitiven Akte, die Wahrnehmung und Handeln in unseren alltäglichen Aktionen konstituieren. Sie umfaßt Bruchteile von Sekunden. Es ist diese letzte Zeitskala, wie wir versuchen werden zu zeigen, auf der wir innere Repräsentationen lokalisieren können.

Wie wir gesehen haben muß ein kognitives System durch Selbstorganisation dazu befähigt werden, gewisse Veränderungen in seiner Umgebung korrekt vorherzusagen, um verhaltensrelevante Entscheidungen zu fällen. Die Einsicht, daß kognitive Systeme erst in einer sensomotorischen Schleife solche Fähigkeiten entwickeln und auch vermöge von Selbstorganisationsprozessen in einer solchen agieren, hat in den letzten Jahren zu einer veränderten Sehweise von Kognition geführt in der dynamische Aspekte eine immer größere Rolle spielen. Umwelt und kognitives System werden dabei als sich wechselseitig spezifizierend angesehen, wobei zu berücksichtigen ist, daß zur "Umwelt" im allgemeinen auch andere kognitive Systeme gehören. Im Kontext solch "situierter" Systeme werden innere Repräsentationen nicht mehr als pure Reflexion der externen Welt verstanden, und bisweilen wird der Begriff der "inneren Repräsentation" - als obsolet - vollständig abgetan (Brooks 1991), da er, wie wir zeigen werden, nicht mehr auf (speicherbare) materielle oder formale Strukturen zurückzuführen ist. Repräsentation hat zwar ihren Ursprung in der physikalischen Außenwelt und in der Gegebenheit des biologischen Körpers, aber sie wird erst in dem Moment dynamisch "internalisiert", in dem interne kognitive Prozesse die Koordination der verschiedensten lokalen Teildynamiken übernehmen und damit letztlich adäquates Verhalten bewirken. Die beteiligten Strukturelemente (z.B. Neuromodule in dem unten beschriebenen Ansatz) können aber bei veränderten inneren oder äußeren Situationen vollkommen andere Funktionen übernehmen. Außerdem kann ein und dieselbe Funktion von verschiedensten Strukturelementen realisiert werden. Innere Repräsentation verliert damit aber ihre (persistente) systemintrinsische Bedeutung, sie existiert nur für kurze Zeit als Teilprozeß in einer kognitiven Dynamik.

Unsere systemtheoretische Betrachtung kognitiver Systeme akzeptiert das Paradigma der neuronalen Netze (s.u.) als Grundlage der Beschreibung ihres materiellen Substrats. Kanonisch ist dann die Annahme, daß das Verhalten, d.h. die Veränderung der neuronalen Aktivität, eines kognitiven Systems im Prinzip durch ein zu spezifizierendes

nichtlineares dynamisches System, d.h. ein System von gekoppelten nichtlinearen Differentialgleichungen erster Ordnung, zu beschreiben ist. Kognitive Prozesse sind als globale, verteilte Prozesse zu verstehen, die das Ergebnis vieler kooperativ oder konkurrierend miteinander wechselwirkender Moduldynamiken sind. Sie werden in Aufnahmen der neuen bildgebenden Technologien als beständig sich verändernde, breitgefächerte Muster neuronaler Hirnaktivität "sichtbar". Der dynamische Ansatz zur Beschreibung kognitiver Systeme ist also nicht daran orientiert, die Mechanismen einer etwaigen Informationsverarbeitung in Gehirnen aufzudecken, sondern vielmehr die Emergenz verhaltensrelevanter Aktivitätsmuster des Gehirns im Rahmen eines neuronalen Selbstorganisationsprozesses zu verstehen.

3 Grundbegriffe der Theorie dynamischer Systeme

Bevor wir daran gehen, kognitive Systeme als modulare neurodynamische Systeme genauer zu charakterisieren, werden wir einige der benötigten Grundbegriffe der Theorie dynamischer Systeme einführen (vgl. Wiggins 1990; Jackson 1991; Abraham und Shaw 1992; Ott 1993). Ausgangspunkt einer systemtheoretischen Beschreibung ist der Begriff des *Zustands*. Darunter verstehen wir einen Satz von Systemgrößen, der die zeitliche Entwicklung des Systems vollständig zu beschreiben erlaubt. Für ein klassisches physikalisches Punktteilchen sind dies z.B. der Ort und die Geschwindigkeit des Teilchens. In unserem Fall kann man dabei an die Aktivitäten (das Somapotential am Axonhügel) der Zellen und an die Stärke der Synapsen denken. Die Menge aller Zustände, die das System einnehmen kann bezeichnen wir als den *Zustandsraum* (oder Phasenraum) M des Systems. Zu einer gegebenen *Anfangsbedingung* $x(t_0)$ (Zustand x zur Anfangszeit t_0) ist die Entwicklung des Systems dann durch eine kontinuierliche oder diskrete zeitliche Folge von Zuständen zu beschreiben. Eine solche Zustandsfolge, die eindeutig durch eine Anfangsbedingung gegeben ist, heißt ein *Orbit* des Systems. Im kontinuierlichen Fall ist dies die Lösungskurve eines Differentialgleichungssystems erster Ordnung. Die Menge aller Lösungskurven heißt Fluß- oder *Phasendiagramm* des Systems. Es beschreibt das Verhalten eines Systems zu allen möglichen Anfangsbedingungen. Ein Zustandsraum zusammen mit einem solchen Phasendiagramm (oder äquivalent: mit einem entsprechenden Differentialgleichungssystem) nennen wir ein kontinuierliches dynamisches System oder kurz eine *kontinuierliche Dynamik*.

Die Orbits eines diskreten dynamischen Systems sind durch die iterierte Anwendung einer Abbildung F von M auf sich selbst gegeben, d.h. es ist $x(t+1) = F(x(t))$. Entsprechend heißt die Abbildung $F : M \rightarrow M$ auch *diskrete Dynamik*. Zu jeder kontinuierlichen Dynamik läßt sich in der Regel auch eine diskrete Dynamik (formal) angeben, z.B. durch die "stroboskopische" Beleuchtung eines kontinuierlichen Orbits oder durch die Markierung der zeitlichen Folge von Durchstoßpunkten eines kontinuierlichen Orbits durch eine geeignet gewählte Ebene im Zustandsraum.

Umgekehrt jedoch ist einer diskreten Dynamik eine kontinuierliche Dynamik nicht eindeutig zuzuordnen. Geht es, wie in unserem Fall, darum, auf qualitativer Ebene das Verhalten eines Systems *im Prinzip* zu verstehen, so ist der Bezug auf eine entsprechende diskrete Dynamik in keiner Weise mit Einschränkungen verbunden.

Wir unterscheiden konservative von dissipativen Systemen. In *konservativen* Systemen existieren sogenannte Erhaltungsgrößen, z.B. die Gesamtenergie in einem klassischen physikalischen System. In einem *dissipativen* System gibt es dagegen einen Austausch, z.B. von Energie mit der Umgebung. Gehirne, oder hier besser neuronale Netze, verstehen wir entsprechend als nichtlineare dissipative Systeme. Entscheidend für die Komplexität der Dynamik ist die Nichtlinearität, die hier durch die entsprechenden Eigenschaften der Neuronen ins Spiel kommt.

Globale (qualitative) Eigenschaften einer Dynamik werden durch charakteristische Elemente beschrieben. Bei den hier interessierenden dissipativen Systemen sind dies z.B. deren Attraktoren. *Attraktoren* sind Mengen von Zuständen, gegen die bei fortschreitender Zeit alle Orbits in deren Umgebung konvergieren, d.h. Attraktoren ziehen quasi die Zustände in ihrer Umgebung an. Wir unterscheiden *Fixpunktattraktoren* und *nicht-triviale Attraktoren*. Ein Fixpunkt ist ein stationärer Zustand, also einer, der sich zeitlich nicht verändert (auch trivialer Orbit genannt). Unter den nicht-trivialen Attraktoren finden wir *periodische Orbits* wie den sogenannten Grenzkreisattraktor, d.h. Orbits die nach einer endlichen Zeit T ihren Anfangszustand (und damit auch alle ihre anderen Zustände) wieder durchlaufen, d.h. es gilt $x(t+T)=x(t)$ für alle Zeiten t. Die kleinste solche Zeit T heißt die *Periode* des Attraktors; er beschreibt ein T-periodisches (oszillierendes) Systemverhalten. Attraktoren können aber auch höherdimensionale Gebilde sein, wie die sogenannten Tori, auf denen *quasiperiodische Orbits* dicht liegen. *Chaotische Attraktoren* sind u.a. dadurch charakterisiert, daß sie fraktale Mengen sind. Sie tragen die *chaotischen Orbits*, die auf ihnen dicht liegen (vgl. z.B. Jackson 1991; Abraham und Shaw 1992, Ott 1993).

Orbits, die auf einen Attraktor "zufließen", bezeichnen wir auch als *Transienten* eines Systems. Die Zeit, die vergeht bis das System von einem Anfangszustand entlang der Transienten in die nähe des Attraktors kommt, kann unter Umständen sehr lange währen. Die Existenz nicht-trivialer Attraktoren führt oft zu raum-zeitlichen Mustern, den *dissipativen Strukturen*. Ihr Auftreten ist eng mit dem Begriff der Selbstorganisation eines Systems verknüpft. Das "Erscheinen" eines nicht-trivialen Attraktors wird dann auch als *emergente* Struktur- oder Musterbildung bezeichnet (vgl. z.B. Kelso 1995).

Ein System kann mehrere Attraktoren gleichzeitig besitzen; wir sprechen dann von *koexistenten Attraktoren*. In diesem Fall sind die Bassins der Attraktoren interessant: Das *Bassin* eines Attraktors ist die Menge aller Zustände, deren Orbits mit fortschreitender Zeit zu diesem Attraktor fließen, also quasi sein "Einzugsgebiet". Die

Grenzen zwischen verschiedenen Bassins können sehr "irreguläre" Gebilde, z.B. fraktale Mengen sein. Auf welchem Attraktor die Dynamik des Systems letztlich "endet" hängt von der Wahl der Anfangsbedingungen ab. In Abb. 5b werden z.B. die Bassins von zwei koexistenten Attraktoren eines Neuromoduls gezeigt. Im folgenden wollen wir ein System (genauer: das Verhalten eines Systems) *konvergent* nennen, wenn seine Dynamik nur Fixpunktattraktoren besitzt. Existiert in der Menge der Attraktoren mindestens ein periodischer (quasiperiodischer, chaotischer) Attraktor, so heißt das System *periodisch* (*quasiperiodisch*, *chaotisch*).

Ein nichtlineares dissipatives System kann sich also, abhängig von den Anfangsbedingungen, qualitativ auf verschiedene Weise verhalten, wenn es verschiedene koexistente Attraktoren besitzt. Die zugehörige Basinstruktur ist für konvergente Systeme relativ regulär. Insbesondere beim Vorliegen von chaotischen Attraktoren sind die Bassins aber oft dicht ineinander verwoben, besitzen sogenannte fraktale Grenzen, so daß es sehr schwierig ist bei ungenauer Kontrolle der Anfangsbedingungen das zukünftige asymptotische Verhalten, d.h. den wirksamen Attraktor, des Systems vorherzusagen (vgl. Abb. 5b). Andererseits ermöglicht diese Eigenschaft aber den Wechsel von einem Bassin in ein anderes schon durch das Anlegen einer kleinen Störung. Daß dies schon in einfachen Neuromodulen zu realisieren ist, werden wir weiter unten demonstrieren.

Komplizierter wird die Situation aber noch dadurch, daß die Dynamik von sogenannten *Kontrollparametern* abhängig sein kann. Dies sind Größen, die sich relativ zur "inneren" Dynamik des Systems nur langsam verändern. Grob gesprochen bedeutet dies: Bevor sich ein Kontrollparameter ändert, muß das System Gelegenheit haben, sich einem Attraktor zu nähern. Haben wir bei der Beschreibung eines Systemverhaltens solche Kontrollparameter zu berücksichtigen, dann ist das System durch eine *parametrisierte Familie* von Dynamiken zu beschreiben. Also durch eine Abbildung $F = F(x;\mu)$, $x \in M$, wobei μ einen Satz (Vektor) von Kontrollparametern bezeichnet. Abhängig von den Kontrollparametern kann sich das Verhalten des Systems qualitativ ändern: Z.B. kann ein Fixpunktattraktor verschwinden und statt dessen ein periodischer Orbit erscheinen. Kontrollparameterwerte bei denen eine solche qualitative Verhaltensänderung stattfindet heißen kritische oder *Bifurkationspunkte*. Eine solche Folge von "Verhaltenssprüngen" kann in *Bifurkationsdiagrammen*, wie dem in Abb. 4b, verdeutlicht werden: Aufgetragen wird der zu einem Parameterwert existierende Attraktor. Gut zu erkennen ist in Abb. 4b, mit wachsendem Kontrollparameter *I*, eine Folge von Attraktoren der verschiedensten Typen: zunächst Periode-2 (zwei Punkte), dann Fixpunkte (ein Punkt), quasiperiodische Attraktoren (dichte Punktmengen) und weiter Attraktoren mit höheren Perioden (p Punkte), chaotische Attraktoren (dichte Punktmengen) und so weiter. (Um quasiperiodische und chaotische Attraktoren zu unterscheiden ist es oft notwendig, ihre sogenannten *Liapunov Exponenten* zu berechnen.)

Betrachten wir die Dynamik zweier Teilsysteme die durch zwei Attraktoren des gleichen Typs charakterisiert sein möge, d.h. beide Attraktoren sind periodisch mit Periode *T*, quasiperiodisch oder chaotisch. Verlaufen die Orbits der entsprechenden Teilsysteme *asymptotisch*, d.h. für $t \to \infty$, in einer festen Phasenbeziehung, so nennen wir die Dynamik der Teilsysteme *kohärent*. Ist im Spezialfall die Phasendifferenz gleich Null, so verläuft die Dynamik der Teilsysteme *synchron*. Die Kohärenz bzw. Synchronizität der Dynamik von Teilsystemen wird bei der Beschreibung kognitiver Prozesse eine wichtige Rolle spielen.

Mit Bezug auf die Dynamik der hier betrachteten kognitiven Systeme wollen wir den Begriff des *semantischen Bereichs* eines Teilsystems wie folgt einführen. Da sich einerseits die Eingangssignale zu kognitiven Systemen wegen der sich ständig wandelnden Umweltsituationen ändern, und andererseits die innere Aktivität des Systems wegen der vielfachen Kopplungen Einfluß auf jedes Teilsystem nimmt, wird die Dynamik eines solchen Teilsystems meist in der Nähe eines Attraktors verlaufen und selten auf ihn gelangen; d.h. Neurodynamik ist im wesentlichen eine Transientendynamik. Wir werden der Dynamik eines Teilsystems "Bedeutung" dann zuschreiben, wenn durch sie eine verhaltensrelevante Leistung des Systems ermöglicht wird, und wir gehen davon aus, daß eine Transientendynamik, die im gleichen Bassin des Teilsystems verläuft, letztlich die gleiche verhaltensrelevante Leistung bewirkt und damit die gleiche Bedeutung trägt. "Bedeutungsträger" ist also das Bassin eines Attraktors, d.h. die Menge aller Zustände, die durch den Attraktor eindeutig (s.o.) charakterisiert ist. Der *semantische Bereich* eines Teilsystems ist dann durch das Bassin eines Attraktors definiert, dessen Transienten für die Realisierung eines adäquaten Verhaltens alle die gleiche Bedeutung besitzen. Nach dieser Definition kann (aber muß nicht) jeder Attraktor einen semantischen Bereich des Teilsystems charakterisieren. Ob ein dynamischer Bereich "semantisch" ist kann allerdings erst nach Registrierung und Bewertung der entsprechenden Wirkung entschieden werden. Semantische Bereiche eines Teilsystems sind also kein systemintrinsisches Charakteristikum.

4 Neuronale Systeme

Als "neuronale Netze" werden heute sowohl technische Anwendungen in Form funktional beschränkter künstlicher neuronaler Netze als auch die "realistische" Modellierung biologischer Gehirne bezeichnet. Wir verwenden den Begriff der neuronalen Systeme hier im Sinne des modernen Konnektionismus als Paradigma zur Beschreibung des Gehirns bzw. hirnähnlicher Strukturen. Anders als die Künstliche Intelligenz (KI) bzw. die KI-dominierten Kognitionswissenschaften verweist der parallel entstandene Konnektionismus auf das den betrachteten Phänomenen zugrunde liegende neuronale Substrat. Im Unterschied zur Kybernetik (der ersten Art), die die Adaptations- und Lernfähigkeit dieser Systeme schon als dynamische Selbstorganisationsprozesse begriff (Yovits und Cameron, 1960; von Foerster und Zopf,

1962), vermag er aufgrund der inzwischen vollzogenen mathematischen Entwicklung diese Vorstellungen präziser zu inkorporieren.

Das Paradigma der neuronalen Netze geht von den folgenden idealisierenden Hypothesen aus: Die zu studierenden, insbesondere kognitiven Fähigkeiten und Funktionen von Gehirnen beruhen im wesentlichen auf der komplexen Verknüpfungsstruktur einer großen Zahl relativ einfacher nichtlinearer Elemente, den Neuronen. Die spezifischen biochemischen und biophysikalischen Eigenschaften des einzelnen Neurons, mit Ausnahme der Nichtlinearität, sind in bezug auf diese Fähigkeiten von nachgeordneter Bedeutung. Die Wirkung eines Neurons auf ein anderes wird allein durch Aussendung von Signalen entlang des Axons vermittelt und hängt von der Stärke entsprechender Synapsen ab. Auf diesem Hintergrund werden die Neuronen als Punktzellen verstanden, d.h. die räumlichen Integrationseigenschaften und nichtlinearen Leitungseigenschaften der Dendritenbäume werden vernachlässigt. Weiterhin werden die räumliche Anordnung von Synapsen und die Verzweigungsformen der dendritischen und axonalen Strukturen als für die operationalen Eigenschaften eines Netzwerkes unwesentlich erachtet. Der Einfluß des chemischen Umfelds von Neuronen sowie deren komplexe interne chemische und molekularbiologische "Maschinerie" bleibt unberücksichtigt.

Als erster Schritt auf dem Weg zum Verständnis funktionaler Eigenschaften biologischer Gehirne erscheint diese Abstraktion durchaus sinnvoll. Außerdem ist die Überprüfung von Hypothesen bezüglich der Prinzipien neuronaler Signalverarbeitung an solchen *künstlichen neuronalen Strukturen* nicht nur sinnvoll, sondern auch notwendig, z.B. wegen der in biologischen Systemen kaum zu kontrollierenden Randbedingungen. Insbesondere geht es bei Untersuchungen an künstlichen Systemen dieser Art um den Zusammenhang von Struktur und Funktion, d.h. um die Rolle der Netzwerkarchitektur (Verküpfungsstruktur) bei der Realisierung eines bestimmten Verhaltensrepertoires bzw. bestimmter Fähigkeiten, um die Bedeutung externer Randbedingungen für die evolutionäre Gehirn- bzw. Strukturentwicklung sowie um das Verständnis der inneren Mechanismen, die den Adaptations- und Lernprozessen zugrunde liegen.

Künstliche neuronale Systeme gehen von sehr einfachen Elementen aus, den sogenannten *formalen Neuronen*. Diese summieren in der Regel die (durch Synapsenstärken) gewichteten Eingangssignale einer Zelle um die Aktivität (das Membranpotential) zu bilden. Das Ausgangssignal wird dann mittels einer nichtlinearen Funktion in Abhängigkeit von der Aktivität erzeugt, bisweilen (und näher an der Biologie) auch in Form eines Aktionspotentials ("Spikes") in Abhängigkeit von einem Schwellwert. Zu Demonstrationszwecken werden wir weiter unten Beispiele künstlicher neuronaler Netze wählen, die auf dem einfachsten nichtlinearen Neuronenmodell beruhen. Dies ist wie folgt gegeben:

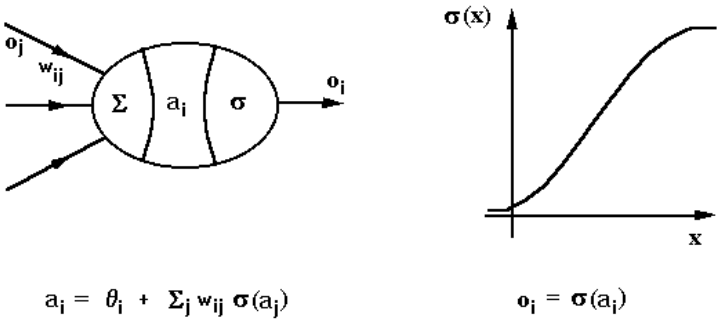


Abb. 2: Ein formales Neuron und seine Transferfunktion σ.

Dabei bezeichnet a_i die Aktivität, θ_i den Schwellwert und $o_i = \sigma(a_i)$ den Ausgang des Neurons i. Die nichtlineare Funktion $\sigma(x) := (1 + e^{-x})^{-1}$ wird *Transferfunktion* des Neurons genannt, und w_{ij} kennzeichnet die Synapsenstärke der Verbindung von Neuron j zu Neuron i. Realistischere Modelle biologischer Neuronen, wie das Hodgkin-Huxley Modell (Hodgkin und Huxley 1952) oder das FitzHugh-Nagumo Modell (FitzHugh 1969; Nagumo et.al. 1962), sind bereits so komplex, daß die Simulation größerer Netzwerke aus solchen Elementen die vorhandenen Computerressourcen oft übersteigt. Biologienähere formale Neuronen sind z.B. die "spikenden" Neuronen (vgl. Gerstner et.al. 1993) und ein von Aihara et.al. (1990) entwickeltes Modell mit komplexeren Eigenschaften.

Die Emergenz kognitiver Fähigkeiten setzt sicher einen kritischen (hohen) Grad an Komplexität des unterliegenden Systems voraus. Dennoch ist es sinnvoll das Verhalten bzw. die funktionalen Eigenschaften kleiner neuronaler Strukturen zu studieren. Solche meist funktional oder räumlich beschränkte Netzwerke werden im folgenden *Neuromodule* genannt. Dies verweist darauf, daß sie als Subsysteme eines größeren neuronalen Systems gedacht werden. Da das Modulkonzept für die folgenden Betrachtungen grundlegend ist, werden wir uns kurz einigen formalen Aspekten dieses Begriffs zuwenden.

4.1 Neuromodule

Ein formales Neuromodul ist zunächst gekennzeichnet durch die Zahl und den Typ seiner Neuronen sowie durch seine Architektur. Wir unterscheiden zwischen einer *rekursiven* Architektur, in der geschlossene, gerichtete Schleifen von Verbindungen existieren, und einer schleifenlosen *gerichteten* Architektur. (Als *rekurrent* bezeichnet man heute meist Netzwerke, deren Ausgänge auf die Eingänge zurückkoppeln. Wir betrachten sie hier als Spezialfall rekursiver Architekturen.) Rekursive Architekturen ermöglichen unter bestimmten Bedingungen eine komplexe Moduldynamik, gerichtete Architekturen, wie die der feedforward Netze (Minski und Papert, 1969; Hertz et.al.

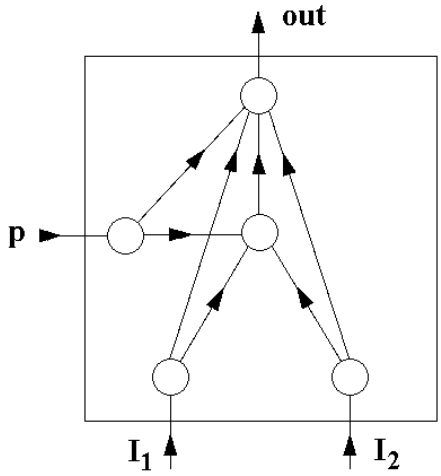
1991), sind dynamisch trivial, d.h. sie sind konvergent (s.u.). Wichtig für unsere Betrachtungen sind also primär Module mit einer rekursiven Architektur.

Zunächst gehen wir von der Vorstellung aus, daß ein solches Modul im Gesamtverband der Module eines Systems eine gewisse Funktion ausübt, d.h. eine funktionale Einheit bildet. Um diese Vorstellung zu präzisieren müssen wir *Schnittstellen* definieren. Dazu unterscheiden wir Signaleingänge und -ausgänge sowie Parametereingänge und -ausgänge des Moduls. Die Signaleingänge bzw. -ausgänge sind Verbindungen, die zu *Eingangsneuronen* hin bzw. von *Ausgangsneuronen* weg führen. Neuronen, die keine solche Verbindungen besitzen werden entsprechend *innere Neuronen* genannt.

Die Funktion eines Moduls bezieht sich auf den Zusammenhang zwischen Ein- und Ausgangssignalen. Zu einem stationärem Eingangssignal kann sich, eventuell nach einer gewissen Übergangszeit (Transienten), z.B. ein stationäres Ausgangssignal einstellen. Ergibt sich zu jedem möglichen stationärem Eingangssignal ein stationäres Ausgangssignal, so heißt das Modul *konvergent*. Die Funktion eines solchen Moduls kann z.B. darin bestehen, Eingangssignale zu klassifizieren bzw. zu kategorisieren oder, allgemeiner, als ein Assoziativspeicher zu wirken (Minski und Papert, 1969; Hertz et.al. 1991). Mathematisch läßt sich das Verhalten eines konvergenten Moduls durch eine nichtlineare Input-Output-Abbildung darstellen. Zu einem stationärem Eingangssignal kann sich aber auch ein periodisches oder gar chaotisches Ausgangssignal einstellen. Wir sprechen dann von einem dynamischen, und genauer, von einem periodischen bzw. chaotischem Verhalten des Moduls. Die funktionale Bedeutung solch dynamischer "Ausgangsmuster" ist, im Gegensatz zu den stationären Ausgängen konvergenter Module, nicht einfach zu bestimmen; es sei denn, die Ausgangssignale steuern Motorneuronen direkt an, so daß ihre Wirkung offensichtlich wird. In der Regel wird man ihre Funktion nur dann erkennen, wenn man die Wirkung solcher Signale in einem Verband von miteinander wechselwirkenden Modulen versteht.

Bedingungen für das Auftreten eines spezifischen Modulverhaltens sind durch die *inneren Parameter* gegeben. Dies können z. B. die Synapsenstärken der Verbindungen und die Schwellwerte der Neuronen des Moduls sein. Innere Parameter können durch Signale an den Parametereingängen verändert werden. Signale dieser Art werden im Modul also nicht "verarbeitet", sondern dienen der Einstellung verschiedener *operationaler Modi* des Moduls. In einem Modul können aber auch Signale erzeugt werden, die über bestimmte Verbindungen nur die Parameter anderer Module beeinflussen; wir nennen die entsprechenden Verbindungen Parameterausgänge. Die Existenz von Parametereingängen erlaubt ein multi-funktionales bzw. kontextabhängiges Verhalten des Moduls: Zu ein und demselben externen Eingangssignal können, abhängig vom Kontext, d.h. von Signalen an den Parametereingängen, verschiedene Ausgangssignale erzeugt werden. Dieser Zusammenhang soll an einem sehr einfachen Beispiel verdeutlicht werden:

Bestimmte neuronale Netze sind in der Lage mathematische Funktionen zu realisieren, oder genauer, zu approximieren. Zum Beispiel sind binäre (Boolesche) Funktionen in feedforward Netzen zu implementieren (Minski und Papert, 1969; Hertz et.al. 1991), d.h. ein entsprechend trainiertes Netzwerk ordnet jedem binären Eingangssignal ein definiertes binäres Ausgangssignal zu. Mit einem zusätzlichen Parametereingang ist nun ein konvergentes Modul zu entwickeln, das abhängig vom Signal am Parametereingang verschiedene solcher Funktionen approximiert. Das folgende Modul mit einem inneren Neuron (Abb. 3a) realisiert z.B. zu den Parametereingängen p=0.0, 0.5, 1.0 die Booleschen Funktionen OR, XOR, und NAND (vgl. Abb. 3b); es besitzt also drei operationale Modi.



I_1	I_2	p=0.0 OR	p=0.5 XOR	p=1.0 NAND
0	0	0	0	1
0	1	1	1	1
1	0	1	1	1
1	1	1	0	0

Abb. 3: a.) Ein einfaches multifunktionales Modul approximiert b.) drei Boolesche Funktionen.

An diesem Beispiel ist auch die funktionale Unterscheidung von externen Eingängen und Parametereingängen zu diskutieren. Natürlich ist das beschriebene Modul auch als Modul mit drei Eingängen und einem Ausgang zu verstehen. Die Unterscheidung in externe und Parametereingänge ergibt sich letztlich aus der Funktion, die wir dem Modul zuschreiben, nämlich hier einem zweistelligen binären Eingangssignal kontextabhängig ein einstelliges binäres Signal zuzuordnen. Die Eingänge des Beispiels unterscheiden sich auch dadurch, daß für externe Eingänge nur binäre Werte, als Parameter aber kontinuierliche Werte erlaubt sind.

Parametereingänge können auf verschiedene Weise auf die Modulkonfiguration einwirken. In obigem Beispiel wirkt das Parametersignal *p* additiv, d.h. es wirkt auf die angesprochenen Neuronen vermöge eines zusätzlichen exzitatorischen oder inhibitorischen Signals; es verändert letztlich den Schwellwert der Neuronen. Es sind auch andere Wirkungsformen möglich; z.B. können Parametersignale multiplikativ auf externe Eingangssignale vermöge sogenannter "Synapsen höherer Ordnung" (vgl. Baird 1990) wirken oder die Synapsenstärken direkt verändern.
Das vorgestellte Modulkonzept besitzt drei Arten von Schwierigkeiten. Geht man zunächst von einem Gesamtsystem (wie dem Gehirn) aus, so wird seine Einteilung in

Module nicht sofort offensichtlich sein. Sie wird in der Regel davon abhängen, welche Unterfunktionen (z.B. Beiträge zur visuellen, akustischen, olfaktorischen Wahrnehmung, zum Sprachverständnis und zur Spracherzeugung, zur Aufmerksamkeitssteuerung etc.) man vermutet bzw. ausfindig macht. Setzt man die Multifunktionalität der Module voraus und glaubt man wie wir daran, daß sich ihr spezifisches Verhalten, und damit letztlich auch ihre Funktion, erst während der kooperativen oder konkurrierenden Interaktion mit anderen Modulen ergibt, so ist eine eindeutige Zerlegung des Systems in funktional bestimmte Module kaum durchzuführen. Hat man jedoch eine solche funktionale Untergruppe lokalisiert, so ist zweitens die Aufteilung in Signal- und Parametereingänge bzw. Signal- und Parameterausgänge bis zu einem gewissen Grad von dem Vorverständnis des Aufteilenden abhängig. Die Wahl einer geeigneten Schnittstelle kann für das Verständnis des Modulverhaltens durchaus kritisch sein. Auch hier begegnet uns wieder eine zyklische Kausalität: Einerseits spricht man von Struktur und Funktion der von den Schnittstellen begrenzten Bereiche (Module), andererseits ist gerade über die Schnittstellen Funktion und Struktur dieser Bereiche erst zu verstehen (Glünder 1993). Zur dritten Schwierigkeit: Hat man sowohl die Funktionen des Moduls als auch die entsprechenden Ein- und Ausgangsneuronen spezifiziert, so stellt sich die Frage mit welcher inneren Struktur das gewünschte Verhalten erzeugt werden kann. Dies ist in der Regel rein analytisch bzw. technisch kaum zu entscheiden. Die Neurophysiologie kann bislang nur begrenzt Auskunft über die Verknüpfungsstruktur von Neuromodulen und die Verwendung von exzitatorischen bzw. inhibitorischen Verbindungen geben. Im Bereich der künstlichen neuronalen Netze hat man daher nach Lernregeln gesucht, die zu einer Klasse von Verhalten (in der Regel konvergentes Verhalten) eine geeignete Struktur erzeugen (Hertz et.al. 1991). Zur Erzeugung dynamischen multi-funktionalen Verhaltens sind effiziente Lernregeln noch kaum bekannt. Sicher ist jedoch, daß es immer sehr viele verschiedene Strukturen gibt, die fähig sind, einen bestimmten Satz von operationalen Modi bereitzustellen.

Die Auswahl formaler Neuromodule ist daher, wie vermutlich auch ihre biologische Ausprägung, weitgehend strukturell-ökonomisch begründet. Für analytische (oder technische) Zwecke ist es bisweilen sinnvoll nach minimalen Strukturen zu suchen, die sich in bezug auf die zu realisierenden Funktionen über die zahlenmäßige Beschränkung der inneren Verbindungen (und damit indirekt auch der inneren Neuronen) definieren. Module können sich vermöge geeigneter Verbindungen aus solchen minimalen Strukturen zusammensetzen; diese können zu Meta-Modulen agglomerieren und so fort. Module können also nahezu beliebigen Umfang haben. Wir wollen hier darauf verzichten, spezifische Gehirnstrukturen als Module zu benennen: dies könnten sehr kleine Strukturen sein, aber auch corticale Mikro- oder Hyperkolumnen oder spezifische Areale.

Auf die Bedeutung eines biologischen Modulkonzepts für die funktionale Interpretation der Cortex-Architektur hat z.B. Szentágothai (1983) mehrfach hingewiesen. Vorstellungen vom modularen Aufbau des Gehirns, in dem Sinne, daß es als Netzwerk

zu interpretieren ist, dessen konstititutive Elemente selbst wieder Netzwerke sind, gehen zurück auf Hebbs (1949) "cell assemblies" und auf Freemans (1975) "Katchalsky sets". Modularität ist aber sicher auch ein effektives Designkonzept für eine synthetische Hirnmodellierung und wurde als solches z.B. von Reeke und Sporns (1993) angewandt. Von der oben ausgeführten Möglichkeit einer parametrisierten Moduldynamik, d.h. der Einstellung verschiedener operationaler Modi eines Moduls durch äußere Signale, wurde in theoretischen Betrachtungen bislang nicht explizit Gebrauch gemacht. Sie erlaubt jedoch die Einführung von Multifunktionalität und funktionaler Variabilität schon auf der Ebene der Elemente eines kognitiven Systems und ist damit Grundlage unserer Vorstellungen von der Entstehung innerer Repräsentationen und kognitiver Prozesse in Gehirnen.

4.2 Formale Neurodynamik

Um die Dynamik eines neuronalen Netzwerkes bzw. eines Neuromoduls mit n Neuronen formal zu beschreiben, ist zunächst festzuhalten, daß wir den Zustand des Systems x(t) zur Zeit t durch die neuronalen *Aktivitäten* und die Stärke der Synapsen (die synaptischen *Gewichte*) charakterisieren, d.h. durch den Aktivitätenvektor $a(t) := (a_1(t),...,a_i(t),...,a_n(t))$ und den Gewichtsvektor (bzw. die Gewichtsmatrix) $w(t) := (w_{11}(t),...,w_{ij}(t),...,w_{nn}(t))$ zur Zeit t. Der Zustandsraum $M = A \times W$ eines neuronalen Systems ist also das Produkt aus dem *Aktivitätenraum* A und dem *Gewichtsraum* W, d.h. M ist in der Regel sehr hochdimensional, $dim\ M \approx n(1+n)$. Das aktuelle Verhalten eines Moduls, als Ausdruck der zeitlichen Änderung von Neuronenaktivitäten, wird durch die *Aktivitätendynamik* auf A beschrieben. Die Änderung der Synapsenstärken des Systems, die für die Herausbildung spezifischer Funktionen bzw. Fähigkeiten verantwortlich sind, wird durch die *Gewichtsdynamik* auf W beschrieben. Wichtig ist im folgenden die (vereinfachende) Annahme, daß die Änderung der Neuronenaktivitäten im Netzwerk sehr viel schneller erfolgt als die Änderung der Synapsenstärken. Dies erlaubt uns, die Gewichte w_{ij} als Kontrollparameter der Aktivitätendynamik auf A zu verstehen. Es gibt allerdings Hinweise darauf, daß es tatsächlich "schnelle" Synapsen in biologischen Gehirnen gibt, und einige Modelle (von der Malsburg 1981) gehen davon aus, daß solche "modulierenden" Synapsen wichtige Funktionen bei kognitiven Prozessen übernehmen. Als weitere Kontrollparameter der Aktivitätendynamik kommen noch die stationären (oder sich langsam ändernden) externen Eingangssignale des Moduls und spezifische Neuroneneigenschaften wie der Schwellwert in Betracht.

Die Gewichtsdynamik wird im allgemeinen mit Lernprozessen in Gehirnen assoziiert. Als Grundlage biologischer Lernprozesse wird die *Hebbsche Regel* angesehen. Vereinfacht dargestellt besagt sie, daß die Verbindung zwischen zwei Neuronen gestärkt (bzw. geschwächt) wird, wenn beide Neuronen in einem vorgegebenen Zeitfenster gemeinsam aktiv sind (Hebb 1949). Die Hebbsche Lernregel ist also sehr

allgemein formuliert und bedarf der konkreten formalen Ausformung unter Berücksichtigung entsprechender Neuronenmodelle, Übertragungseigenschaften (Synapsen) und Signaleigenschaften. In welcher Form sie zu den hier interessierenden Selbstorganisationsprozessen beiträgt, ist zur Zeit nicht geklärt. Berücksichtigt man die Tatsache, daß bei den oben angesprochenen Adaptations- und Lernprozessen kognitiver Systeme sowohl phylogenetische (Zahl der Neuronen, Modularisierung, grundlegende Verknüpfungsarchitekturen) als auch ontogenetische Aspekte (Variabilität der Synapsenstärken) von Bedeutung sind, so liegt nahe, daß in bezug auf eine biologische Gewichtsdynamik noch ein großer Forschungsbedarf existiert. Die eher "technischen" Lernalgorithmen, die im Bereich künstlicher neuronaler Netze zur Anwendung kommen, beruhen meist auf Gradientenverfahren. Diese sind immer dann einsetzbar, wenn als Antwort auf ein Eingangssignal ein stationärer Zustand des Systems zu erlernen ist (Klassifikation von Eingangssignalen, Mustererkennung, etc.). Das obige Modul (Abb. 3a) wurde z.B. mit einem solchen Verfahren trainiert. Ist in einer gegebenen Situation jedoch eine bestimmte Folge von Zuständen zu erzeugen (wie z.B. ein periodischer Attraktor), und dies wird bei der Herausbildung von kognitiven Prozessen die Regel sein, so sind diese Methoden nur in speziellen Fällen anwendbar.

Um einen Eindruck von den komplexen Verhaltensmöglichkeiten schon kleiner Neuromodule zu erhalten, wählen wir zu Demonstrationszwecken eines der einfachsten möglichen Beispiele: Wir betrachten die diskrete Aktivitätendynamik eines Moduls mit zwei formalen Neuronen des oben angegebenen Typs. Die Verknüpfungsstruktur entnehmen wir der Abb. 4a: Neuron 1 besitzt zwei Rückkopplungsschleifen, eine Selbstrückkopplung w_{11} und eine Rückkopplung w_{12} über Neuron 2. Die parametrisierte Aktivitätendynamik $F(a;\mu), a \in R^2, \mu \in R^7$, dieses Moduls ist gegeben durch die folgende Gleichung:

$$a_i(t+1) := \sum_{j=1}^{2} w_{ij}\sigma(a_j) + \theta_i + I_i, \quad w_{22} = 0, \quad i = 1,2.$$

Parameter dieser diskreten Aktivitätendynamik sind die Gewichte w_{ij}, die Schwellwerte θ_i, sowie die Eingangssignale I_1 und $I := I_2$. Wählen wir Neuron 1 als inhibitorisches Neuron, d.h. w_{11} und w_{21} sind beide negativ und Neuron 2 exzitatorisch, d.h. w_{12} ist positiv, erhalten wir zu bestimmten festen Parameterwerten das in Abb. 4b wiedergegebene Bifurkationsdiagramm, welches das komplexe Verhalten des Moduls in Abhängigkeit von einem Eingangssignal I an Neuron 2 wiedergibt. Die festen Parameter sind $\theta_1 = -1, w_{11} = -16, w_{12} = 8, \theta_2 = 0, w_{21} = -8$.

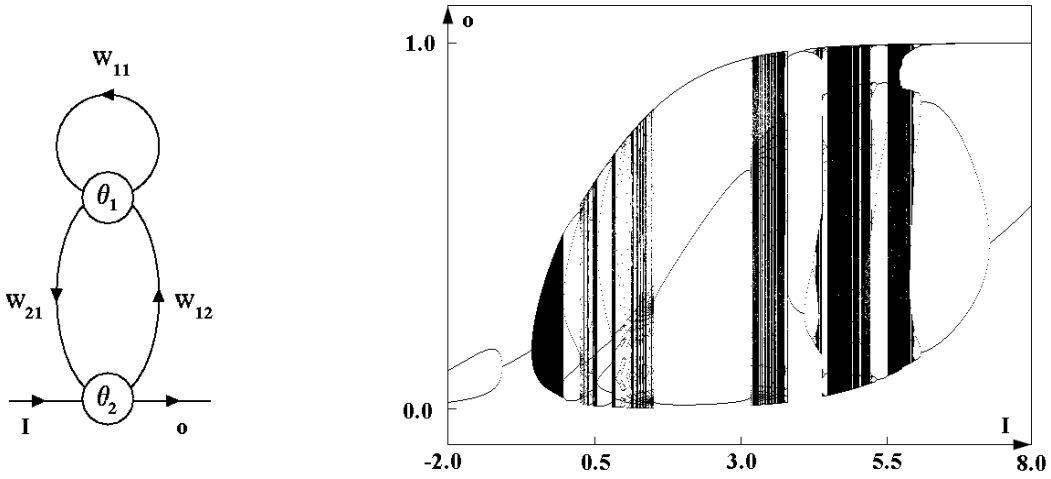


Abb. 4: a.) Ein chaotisches 2-Modul und b.) sein Verhalten in Abhängigkeit vom Eingangssignal I.

Starten wir mit einem negativen Eingangssignal I = -2 so beobachten wir ein oszillierendes Verhalten (Periode 2) des Moduls. Wird der Wert von I langsam erhöht, so erreichen wir ein Intervall mit Fixpunktattraktoren auf das ein Bereich mit quasiperiodischen, dann Periode-5 Attraktoren folgt. Dann schließen sich Attraktoren mit höheren Perioden und chaotische Attraktoren an bis ein größerer Bereich mit Periode-3 Attraktoren erreicht wird. Auf diesen folgen wieder chaotische und periodische Attraktoren bis bei I = 8 wieder Oszillationen mit der Periode zwei beobachtet werden.

In diesem sehr einfachen Neuromodul beobachten wir bereits die Koexistenz von verschiedenen periodischen Attraktoren mit anderen periodischen und chaotischen Attraktoren. Als Beispiel sind in Abb. 5a ein Periode-2 und ein chaotischer Attraktor gezeigt, die für die Schwellwerte $\theta_1 = -0.44$, $\theta_2 = 3.99$ koexistieren. In Abb. 5b sind die zugehörigen Bassins aufgezeichnet, und die "irreguläre" Zerlegung des Phasenraums in diese beiden Bereiche, ihr ineinander "Verwobensein" wird sehr deutlich. Jede Folge von Zuständen, die in dem Bassin des periodischen Attraktors (schwarzer Bereich) verläuft, beschreibt also einen der zwei möglichen semantischen Zustände des Moduls zu dieser Parameterkonstellation. Wegen der komplexen Basingrenzen wird einsichtig, daß viele nahe beieinanderliegende Anfangsbedingungen zu verschiedenen semantischen Zuständen gehören können.

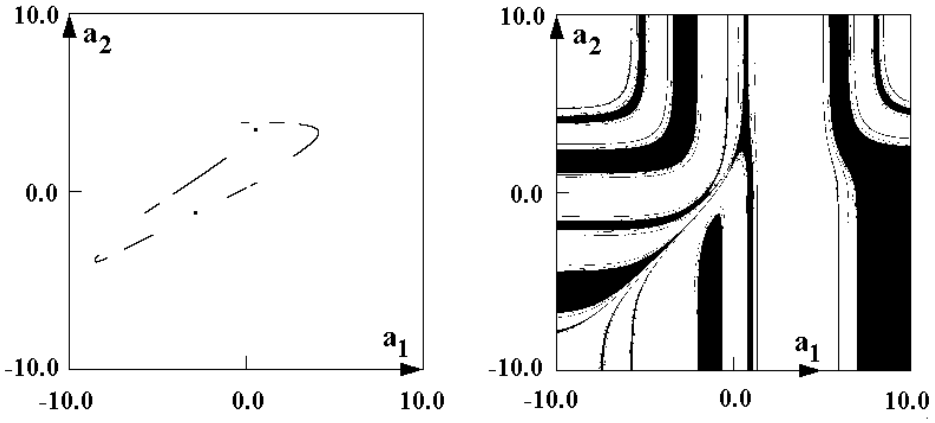


Abb. 5: a.) Ein Periode-2 Attraktor koexistiert mit einem chaotischen Attraktor. b.) Die zugehörigen Bassins (schwarz: Periode zwei).

Neben diesem sehr komplexen dynamischen Verhalten ist insbesondere auch ein dynamischer Effekt interessant, der als *Hysterese* bezeichnet wird. Er kann in solchen Situationen auftreten, in denen zu bestimmten Parameterbereichen verschiedene Attraktoren koexistieren. Im einfachsten Fall beobachten wir ihn bei einem exzitatorischen Neuron mit Selbstrückkopplung $w > 4$ (vgl. Abb. 6a). Hier gibt es einen wohldefinierten Bereich für die Eingangssignale, zu dem zwei Fixpunktattraktoren koexistieren (Bistabilität), außerhalb dieses Bereichs gibt es genau einen Fixpunktattraktor (vgl. Abb. 6b). Starten wir z.B. mit einem stark negativen Eingangssignal und erhöhen dieses langsam, so bleiben wir zunächst in dem Attraktor niedriger Aktivität ($o \approx 0$) bis wir den kritischen Parameterwert θ_1 überschreiten und das System in den Attraktor hoher Aktivität ($o \approx 1$) "springt". Erniedrigen wir nun langsam wieder das Eingangssignal so verbleibt das System zunächst in diesem Attraktor, auch wenn wir den kritischen Punkt θ_1 überschreiten. Das System springt erst im kritischen Parameterwert θ_2 wieder in den Ausgangsattraktor zurück.

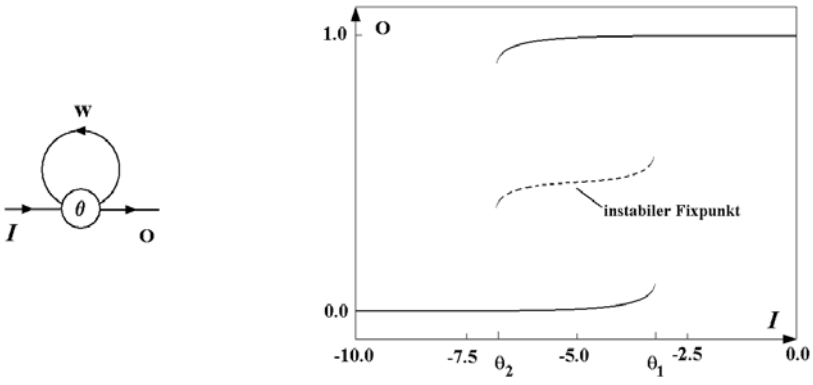


Abb. 6: a.) Neuron mit exzitatorischer Selbstrückkopplung b.) Bistabilität und Hystereseeffekt für w=8.

Dieser Hystereseeffekt kann für alle möglichen Arten von koexistenten Attraktoren auftreten, also auch für periodische und chaotische Attraktoren die mit anderen periodischen oder chaotischen Attraktoren koexistieren. Tatsächlich beobachten wir solche Hystereseeffekte "höherer Ordnung" schon in dem oben diskutierten chaotischen 2-Modul.

Wenn wir uns vorstellen, daß koexistente Attraktoren verschiedene "mentale" Zustände in einem kognitiven neuronalen System charakterisieren könnten, dann wird deutlich, daß die in psychophysikalischen Experimenten beobachteten semantischen, perspektivischen und Figur-Grund-Ambiguitäten (vgl. z.B. Stadler und Kruse 1992) die folgende Interpretation erlauben: Sowohl das "Umkippen" von "Bewußtseinsinhalten" nach ein bis vier Sekunden als auch das "Umkippen" an verschiedenen Stellen einer Folge von solchen mehrdeutigen Bildern kann dadurch erklärt werden, daß zu der jeweiligen Konstellation von Eingangssignalen das entsprechende Teilsystem "bistabil" ist, die jeweiligen Inhalte also mit dem Auftreten von einem der zwei koexistente Attraktoren assoziiert werden. Das "Umkippen" der Inhalte wird im ersten Fall durch ein periodisches "treibendes" Signal, dessen Herkunft zu bestimmen wäre, oder im zweiten Fall durch ein "aufmerksamkeitsgesteuertes" Signal verursacht, welches die Kontrollparameter über den Hysteresebereich des Moduls hinweg bewegt.

Ein weiterer zur Zeit intensiv diskutierter dynamischer Effekt betrifft die "Synchronisation" von oszillierenden Neuronenaktivitäten. In verschiedenen, den olfaktorischen, visuellen und motorischen Cortex betreffenden experimentellen Arbeiten wurde beobachtet, daß Neuronen in entsprechenden Arealen, aber auch in weiter auseinanderliegenden Bereichen, als Antwort auf spezifische externe Stimuli in korrelierter Weise feuern (vgl. Singer 1993; und Artikel in Krüger 1991; Buzsáki et.al. 1994; Pantev et.al. 1995). Es ist zur Zeit nicht offensichtlich, ob dieses "synchrone" Feuern entsprechender Neuronen im Frequenzbereich von 40-70 Hz als eine verhaltensrelevante Eigenschaft oder als ein Epiphänomen zu bewerten ist. Wir wollen hier kurz darauf verweisen, daß die Synchronisation von Neuronenaktivitäten ganz allgemein erfolgen kann, also nicht nur im oszillatorischen Bereich von Neuromodulen, sondern auch beim Vorliegen einer chaotischen Dynamik. Zu Demonstrationszwecken wählen wir zwei der oben beschriebenen chaotischen 2-Module (Abb.4a) und koppeln deren inhibitorische Neuronen auf das exzitatorische Neuron des jeweils anderen Moduls mit der Stärke $w < 0$ zurück. Wir beobachten dann die Synchronisation der beiden Modulaktivitäten über einen weiten Bereich von etwa gleichen Eingangssignalen. Interessant ist, daß zu bestimmten (etwa gleichen) Eingangssignalen zwei verschiedene Attraktoren existieren: Einer der synchronen, ein anderer der asynchronen Modulaktivitäten entspricht (vgl. Abb. 7, mit identischen Modulparametern wie oben angegeben und Modulkopplungen $w = -4$; der asynchrone Periode-2 Attraktor ist durch zwei Punkte gekennzeichnet). In welchem Bassin eine Transientendynamik verlaufen wird, hängt von den Anfangsbedingungen, also von der

"Vorgeschichte" der Module ab. Durch die vorgegebene Kopplung sind bei gleichen Eingangssignalen die Moduldynamiken also kohärent und gegebenenfalls sogar synchronisiert. Wir sagen auch: durch die Kopplung wird die Dynamik der Module *kohärenziert*. Anzumerken bleibt, daß dieser Effekt erhalten bleibt bei kleineren, auch asymmetrischen Variationen der Kopplungsstärken zwischen den Modulen sowie der inneren Modulparameter.

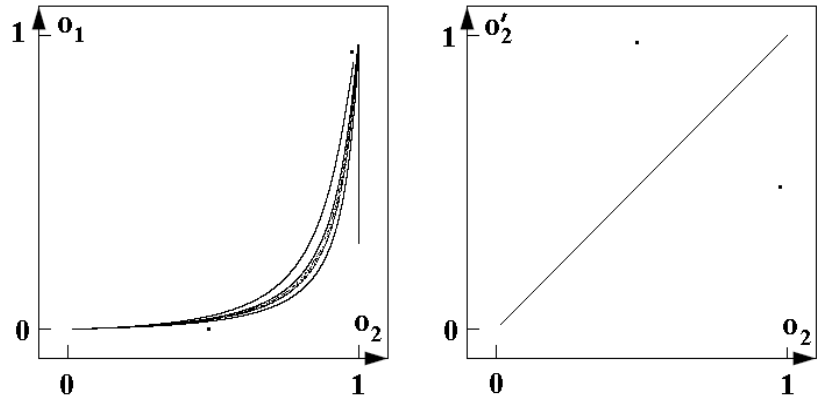


Abb. 7: a.) Koexistenz von Periode-2 und chaotischem Attraktor in gekoppelten Modulen, projiziert auf ein Modul und b.) auf die Ausgänge o_2, o_2' der zwei Module; die chaotische Dynamik verläuft synchron (Hauptdiagonale), die Periode-2 asynchron.

Zum Abschluß sei noch eine weitere Möglichkeit angedeutet, wie mit der komplexen Dynamik schon kleiner neuronaler Strukturen eine Vielzahl von operationalen Modi in einem Modul zu realisieren sind. Aus struktur-ökonomischen Gründen ist es natürlich wünschenswert in einem System mit gegebener Neuronenzahl möglichst viele operationale Modi vorliegen zu haben, die von außen anwählbar sind. Nun ist bekannt, daß in jedem chaotischen Attraktor unendlich viele periodische Orbits existieren, die jedoch alle instabil sind. Wenn es möglich wäre, diese instabilen Orbits zu stabilisieren, dann könnte damit ein im Prinzip unerschöpflicher Vorrat an operationalen Modi erschlossen werden. Tatsächlich wurde von Ott, Grebogi und York (1990) gezeigt, daß ein Kontrollalgorithmus existiert, der dieses leistet. Im Bereich der künstlichen neuronalen Systeme sind inzwischen Neuromodule bekannt, die in der Lage sind, bestimmte instabile periodische Orbits in einem chaotischen Neuromodul zu stabilisieren (Stollenwerk und Pasemann 1996). Durch geeignete Eingangssignale an den Kontrollmodulen sind verschiedene periodische Attraktoren im chaotischen Modul "einstellbar". Als kombiniertes neuronales System, das aus einem chaotischen und einem Kontrollmodul besteht, kann diese Anordnung also eine Reihe von unterschiedlichen oszillierenden Modi einstellen, die für weitere Operationen zur Verfügung stehen.

Die einfachen Beispiele neuronaler "Spielzeugmodelle" dieses Abschnitts sollten zeigen, daß ein sehr komplexes dynamisches Verhalten schon in kleinen neuronalen

Netzen auftreten kann, und zwar dann, wenn mehr als eine geschlossene Signalschleife vorhanden ist und sowohl exzitatorische als auch inhibitorische Verbindungen involviert sind. Um eine ähnlich komplexe *kontinuierliche* (s.o.) Dynamik zu realisieren, bedarf es nur einiger Neuronen mehr. Außerdem ist interessant, daß die Dynamik kleiner Module unter bestimmten Bedingungen das Verhalten sehr großer Netzwerke approximativ wiedergibt, wobei die Aktivität eines Modulneurons der synchronen Aktivität eines ganzen Neuronenpools entspricht (Wennekers und Pasemann, 1996). Simulationen zeigen darüber hinaus, daß auch entsprechende Netzwerke mit biologisch realistischeren, "spikenden" Neuronen ein vergleichbar komplexes Verhalten zeigen. Trotz ihrer extrem vereinfachenden Abstraktion biologischer Strukturen können die angeführten Beispiele formaler neuronaler Systeme einen Eindruck von der möglichen Komplexität einer realen biologischen Neurodynamik vermitteln.

Mit Bezug auf den oben eingeführten Begriff des "semantischen Bereichs" eines Moduls, der mit dem Bassin eines Attraktors eng verknüpft war, zeigen die Beispiele verschiedene Möglichkeiten auf, von einem solchen Bereich in einen anderen zu gelangen. Einmal können die entsprechenden Attraktoren durch Parameteränderungen verschwinden und andere neu entstehen (bezüglich kritischer Parameterwerte vgl. z.B. Abb. 4b). Bei koexistenten Attraktoren zu festen Parameterwerten ist es möglich, die Anfangsbedingungen in einem zeitlich diskreten Akt, z.B. durch ein "gepulstes" Eingangssignal, so neu zu setzen, daß die Dynamik in einem anderen Bassin bzw. semantischen Bereich verläuft. Es können aber auch "Kräfte" oder "Rauschen" die Dynamik über die Basingrenzen hinweg treiben und damit einer neuen "Bedeutung" zuführen. Ist erst die durch Kopplung erzeugte kohärente (synchrone) Dynamik zweier (oder mehrerer) Module im oben ausgeführten Sinne Bedeutung tragend, so sprechen wir von einer *semantischen Konfiguration* kohärenter (synchroner) Moduldynamiken. Durch die Kopplung von Kontrollmodulen mit chaotischen Systemen könnten verschiedene Bedeutungen der Dynamik durch die gezielte Stabilisierung verschiedener periodischer Orbits generiert werden. Wir sehen also, daß auf der dynamischen Ebene eines Systems eine Vielzahl von Möglichkeiten existiert, verhaltensrelevante Leistungen hervorzubringen.

4.3 Biologische Neurodynamik

Seit den 70er Jahren ist die Frage, ob und wie die im vorangegangenen Abschnitt skizzierte Komplexität bei realen Hirnprozessen zu kognitiven Leistungen beiträgt, Gegenstand theoretischer Diskussionen (Harth et.al. 1970; Wilson und Cowan 1972; an der Heiden 1980; Guevara et.al. 1983; Harth 1983; Ermentrout 1984). Die Existenz von vielfältigen oszillierenden (periodischen) Prozessen im Gehirn ist inzwischen durch zahlreiche Experimente belegt. Besondere Aufmerksamkeit richtet sich daher auf den Nachweis von niedrigdimensionalem Chaos in Signalen der Hirnaktivität. Viele

experimentelle Arbeiten deuten auf die Existenz einer chaotischer Dynamik in Gehirnen hin (Skarda und Freeman, 1987; Babloyantz und Destexhe 1986; vgl. auch Elbert et.al. 1994 und Artikel in Duke und Pritschard 1991), wobei die Rolle, die sie bei der Realisierung kognitiver Funktionen spielen kann, nicht eindeutig geklärt ist (Freeman 1992; Tsuda 1992; Babloyantz und Lourenço 1994).

In jüngster Zeit hat sich, gestützt auf zahlreiche Experimente, zunehmend ein Konsens darüber herausgebildet, daß die zeitlich parallele Aktivität von Neuronen eine entscheidende Rolle bei diversen kognitiven Funktionen wie Wahrnehmung, Gedächtnis, Aufmerksamkeit, Bewertung usw. spielt (Singer 1993, 1995; Artikel in Krüger 1991; Buzsáki et.al. 1994; Pantev et.al. 1995). Als neurobiologische Grundlage für diese globalen, verteilten Prozesse hat sich der Begriff der "cell assemblies" bzw. der "neuronalen Ensembles" eingebürgert. Ein solches Ensemble besteht aus Neuronen, die über die verschiedensten, auch weit auseinanderliegenden Cortexareale verteilt sind und während eines begrenzten zeitlichen Bereichs (etwa 200 bis 400 msec) "synchron" feuern. (Der Begriff "synchron" wird in diesem Kontext nicht im exakten physikalischen Sinne benutzt; er besagt vielmehr, daß Neuronen etwa mit der gleichen Frequenz und mit etwa konstanten Phasenbeziehungen feuern, wobei die Phasendifferenzen sehr klein sein können.) Es wird angenommen, daß die Existenz dieser Ensembles starke reziproke Verbindungen zwischen ihren Neuronen voraussetzt. Wie Experimente zeigen kann ein Ensemble offensichtlich von kleineren Untereinheiten seiner Neuronen aktiviert werden, die sowohl mit sensomotorischen als auch mit internen Funktionen in Zusammenhang stehen.

Zahlreiche, meist durch Multielektrodenableitungen gewonnene, experimentelle Ergebnisse vermitteln etwa folgendes Bild: Jedes Neuron eines Cortexareals kann an mehr als einem Prozeß beteiligt sein. Neuronen in ein und demselben kleinen Cortexareal können in vielen verschiedenen Funktionen aktiviert werden. Bei einer vorgegebenen äußeren Bedingung zeigen viele Neuronen den gleichen Erregungszustand; sie werden zu einer "funktionalen Gruppe" zusammengefaßt. Ein Areal kann offenbar mehrere funktionale Gruppen von Neuronen umfassen. Ein Grund für die Existenz dieser zeitlich parallelen Gruppenaktivität kann darin bestehen, daß ein gemeinsames postsynaptisches Neuron so verläßlicher angeregt wird, als durch einzelne Neuronen. Auch bleiben Neuronen mit Verbindungen zu nur einigen wenigen dieser Gruppenneuronen unbeeinflußt; dies erlaubt den "Überlapp" verschiedener funktionaler Gruppen in einem Areal. Die "synchrone" Neuronenaktivität ermöglicht so eine verläßliche Transmission von Signalen zwischen einer Menge von divergent und konvergent verbundenen Neuronen wie z.B. Abeles (1991) in seiner "synfire chain"-Hypothese ausgeführt hat. Wir werden die Existenz funktionaler Gruppen von Neuronen und auch die neuronaler Ensembles als Ausdruck der Aktivität gekoppelter Neuromodule interpretieren.

5 Kognitive Systeme als autotrope Systeme

Im folgenden werden kognitive Systeme als nichtlineare dynamische, neuronale Systeme beschrieben, die sich aus einer Vielzahl miteinander wechselwirkender Module zusammensetzen. Jedes dieser Neuromodule besitzt ein bestimmtes Repertoire an Verhaltensmöglichkeiten, welches mathematisch durch eine parametrisierte Schar von dynamischen Systemen (s.o.) zu charakterisieren ist. Durch die Fixierung geeigneter Parameter kann für kürzere oder längere Zeit eine bestimmte Funktion, ein spezifischer operationaler Modus des Moduls ausgewählt werden. Die Änderung der internen Modulparameter kann durch die Wechselwirkung mit anderen Modulen erfolgen. Die sehr allgemeine Aussage, daß kognitive Systeme ihre Fähigkeiten im Rahmen eines Selbstorganisationsprozesses erlangen, wurde in einigen deskriptiven Modellen (Maturana und Varela 1987; Varela et.al. 1992; Shimizu 1993) sowie in formalisierteren, an der Synergetik orientierten Modellen (Kelso 1995; Haken 1995) versucht zu konkretisieren. Die Entstehung von Selbstorganisationsphänomenen in physikalischen Vielteilchensystemen ist in der Regel qualitativ und bisweilen auch quantitativ nachvollziehbar (vgl. z.B. Haken 1982). Die Selbstorganisation in Form biochemischer und biophysikalischer Prozesse in Gehirnen hat aber die Entstehung kognitiver Fähigkeiten zur Folge; sie ist zielgerichtet und hat ein Überleben sicherndes Verhalten zu bewirken: eine Qualität, die den bislang beschriebenen physikalischen Phänomenen fremd ist.

Im Kontext des modularen Ansatzes stellt sich somit die Frage, wie die verschiedenen Module so "zusammenfinden", daß sie verhaltensrelevante Muster von Neuronenaktivität hervorrufen, d.h. eine globale, verteilte kognitive Dynamik erzeugen. "Wegweiser" zur Kooperation von Modulen, das "tagging" im Sinne von Holland (1995), könnte dabei wieder die Synchronisation von Neuronen sein. Es gibt Hinweise darauf, daß der Grad der Synchronisation von Neuronen die Synapsenstärken der Verbindungen zwischen ihnen kontrolliert (was in etwa der Hebbschen Regel entspricht). Da sowohl sensorseitige Signale als auch interne Prozesse den Grad der Synchronisation von Modulen verändern können, hängt die Stärke der Wechselwirkung zwischen Neuronen bzw. zwischen Modulen von diesen Einflüssen ab. Die "Wegweiser" modularer Kooperation sind also vermutlich zeitlicher und nicht ausschließlich räumlicher Natur.

Wir haben gesehen, daß bei gleichbleibender Verknüpfungsstruktur ein Modul gegebenenfalls eine Vielzahl von operationalen Modi besitzt, die sich durch die Wahl verschiedener koexistenter Attraktoren zu festen Parametern oder durch die Variation der internen Parameter einstellen lassen, und daß diese Veränderung von inneren Parametern durch Signale anderer Module an den Parametereingängen erfolgen kann. Welche dieser vielen inhärenten Funktionen jedoch im Zusammenspiel mit anderen Modulen des Systems "semantisch" werden, d.h. tatsächlich verhaltensrelevante

Leistungen des Systems bewirken, ist erst post hoc, d.h. bei der Beobachtung eines sich "adäquat" verhaltenden Systems zu beurteilen. Als isoliertes System besitzt ein solches Modul nur ein gewisses Potential an Verhaltensmöglichkeiten. Seine definitiven, in einem kognitiven Prozeß wirksam werdenden funktionalen Eigenschaften sind jedoch zunächst unbestimmt. In diesem Sinne nennen wir ein isoliertes Modul *funktional unbestimmt*. Die stärkere Hervorhebung der funktionalen Flexibilität bei gleichbleibender Struktur der Systemelemente ist der Kern unserer Beschreibung kognitiver Systeme als *autotrope Systeme*. Mit dem Begriff "autotrop" wollen wir die spezifische Fähigkeit eines Systems bezeichnen, seinen Selbstorganisationsprozeß mit Hilfe dieser inneren Flexibilität an die sich ständig verändernden Randbedingungen zielgerichtet anzupassen. Selbstorganisation in autotropen neuronalen Systemen beruht also nicht auf der durch eine eindeutig festgelegte Wechselwirkung der Elemente hervorgerufenen Reaktion des Systems auf eine vorgegebene externe Randbedingung, sondern insbesondere auf den vielfältigen Möglichkeiten zur Beeinflussung "innerer" Parameter seiner Elemente, die wiederum über ihre jeweiligen dynamischen Eigenschaften die Stärke der Wechselwirkungen (Synapsenstärken) bestimmen.

Die Grundeigenschaften eines autotropen Systems sind etwa wie folgt zu charakterisieren: Es besteht aus Elementen, die als *isolierte* Einheiten *funktional unbestimmt* sind. Die Elemente erhalten ihre definitiven funktionalen Eigenschaften erst dann, wenn sie mit anderen Elementen des Systems in Wechselwirkung treten. Die Wechselwirkung zwischen den Elementen hängt wiederum von den entstehenden Eigenschaften der beteiligten Elemente ab. Sowohl die Elementeigenschaften als auch die Wechselwirkung zwischen den Elementen bilden sich erst in einem zyklischen Prozeß heraus. Die Eigenschaften des autotropen Systems insgesamt emergieren als globale Form der in diesem Selbstorganisationsprozeß aufscheinenden Wechselwirkungen und Elementeigenschaften. Sie entwickeln sich gemeinsam mit den Eigenschaften der Elemente und ihrer Wechselwirkungen aus einer funktional unbestimmten Anfangssituation heraus (Shimizu 1993).

Erscheint dies zunächst als der bekannte Ablauf eines Selbstorganisationsprozesses, so sei darauf hingewiesen, daß die Beschreibung autotroper Systeme sich von der herkömmlicher physikalischer Systeme dadurch unterscheidet, daß die Eigenschaften ihrer isolierten Elemente unbestimmt sind. Die Grundelemente physikalische Systeme, wie z.B. die Elementarteilchen, zeichnen sich gerade durch die Konstanz ihrer charakteristischen Eigenschaften wie Ladung, Ruhemasse etc. aus, d.h. diese das isolierte (freie) Teilchen definierenden Eigenschaften werden nicht nur in der Wechselwirkung mit anderen Elementarteilchen beibehalten, sie bestimmen die Wechselwirkung auch eindeutig. Dieser Unterschied ist vermutlich der Schlüssel zum Verständnis jener beeindruckenden Lernfähigkeit, die wir an biologischen neuronalen Systemen beobachten.

Auf diesem Hintergrund liegt es nahe als Elemente autotroper neuronaler Systeme solche Neuromodule zu wählen, die befähigt sind, das vollständige Spektrum von

Verhaltensmöglichkeiten bereitzustellen: Vom stationären Verhalten (Fixpunktattraktoren), über periodisches oder quasiperiodisches Verhalten bis hin zum chaotischen Verhalten. Module mit dieser Eigenschaft haben wir chaotische Module genannt und oben (Abb. 4a) eines von vielen möglichen Beispielen im Rahmen künstlicher Netzwerke aufgezeigt. Es sei angemerkt, daß hier nicht davon ausgegangen wird, daß die chaotische Moduldynamik unbedingt in kognitiven Akten benutzt wird; es geht vielmehr um die Bereitstellung eines möglichst großen Verhaltensrepertoires, das ist in der Umgebung chaotischer Attraktoren am größten ist (vgl. Abb. 4b). Festzuhalten bleibt, daß solche Module (also auch die biologischen) immer aus exzitatorischen *und* inhibitorischen Neuronen bestehen, wobei die Rolle der inhibitorischen Verbindungen darin bestehen könnte, das System in der Nähe kritischer Parameterwerte zu halten, da dort die größte Verhaltensvariabilität besteht. Auf diese Weise wäre eine weitgehende funktionale Flexibilität des Systems gewährleistet.

Die emergierenden Eigenschaften kognitiver autotroper Systeme sind jedoch nicht beliebig. Ihre besondere Fähigkeit besteht ja gerade darin, in der Lage zu sein, sich an die Bedingungen einer sich verändernder Umweltsituation anzupassen und sich in ihrer Umwelt "adäquat" zu verhalten. Dies bedeutet aber, daß die sensor- und motorseitige Dynamik, in der sich die zeitliche Veränderung externer Situationen darstellt, entscheidende Randbedingungen für die Entfaltung der intrinsischen Selbstorganisationsdynamik setzt. Einige Aspekte des Zusammenspiels von externen Randbedingungen und intrinsischer Systemdynamik seien genauer untersucht. Betrachten wir zunächst ein einfaches Beispiel: Wir wählen als Elemente zwei chaotische Neuromodule A und B mit fixierten inneren Parametern und stationären externen Eingangssignalen, deren Dynamik durch zwei entsprechende Attraktoren charakterisiert ist. Die Module mögen nun mittels der synaptischen Kopplung geeigneter Modulneuronen miteinander in Wechselwirkung treten. Betrachten wir zunächst nur die Kopplung von A nach B dann wird, je nach Zustand des Moduls A, die Dynamik des Moduls B mehr oder weniger beeinflußt: Zu gleichbleibendem externen Eingangssignal stellt sich eine bestimmte Dynamik des Moduls B ein, die dann z.B. durch einen anderen Typ von Attraktor charakterisiert sein kann. Koppelt nun B auf A zurück, so wird dies die Dynamik von A verändern, was wiederum eine Änderung der Dynamik von B bewirken kann, und so fort. In diesem zirkulären Prozeß kann sich eine Situation einstellen, in der beide Module Attraktoren des gleichen Typs besitzen und die Moduldynamiken kohärent sind. Dies kann sich in der korrelierten oder im Spezialfall auch synchronen Aktivitätenänderung entsprechender Neuronen äußern (vgl. Abb. 7) bzw. im experimentell beobachteten "synchronen" Feuern biologischer Neuronen.

Werden die inneren Modulparameter - z.B. durch Signale anderer Systembereiche an den Parametereingängen - verändert, so kann diese Kohärenz auch bei gleichbleibenden externen Signalen aufgelöst werden: Die Module werden desynchronisiert bzw. "dekohärenziert". Weiterhin wird in einer anderen Situation, z.B. bei einer anderen "inneren Disposition" des Systems, die einer anderen Konstellation von inneren

Modulparametern entspricht, die gleiche Konfiguration von Eingangssignalen nicht zur gleichen kohärenten Moduldynamik führen. Ob sich eine bestimmte Kohärenz von Moduldynamiken zu gegebenen Eingangssignalen einstellt hängt also einerseits von den durch die Parameter gegebenen operationalen Modi ab, in denen sich die Module befinden, andererseits aber auch von der Wechselwirkung, also den synaptischen Kopplungen zwischen den Modulen. Die Dynamik wechselwirkender Module kann also durch externe Eingangssignale, innere Parameter und die Stärke der Modulkopplungen synchronisiert oder desynchronisiert, bzw. allgemeiner, kohärenziert oder dekohärenziert werden.

Betrachten wir nun eine große Zahl gekoppelter Module, so wird sich bei gegebener Eingangssituation nur für einen Teil der Module eine kohärente Dynamik einstellen. Erlangt diese kohärente Dynamik Bedeutung insofern als sie eine verhaltensrelevante Leistung des Systems zur Folge hat, so sprechen wir von einer *semantischen Konfiguration* von Moduldynamiken. Eine semantische Konfiguration ist als erlernte Antwort des Systems auf eine gegebene Eingangssituation zu verstehen. Wir charakterisieren sie durch das Bassin jenes Attraktors, der die kohärente Dynamik des entsprechenden Ensembles von Modulen beschreibt. Der Begriff der semantischen Konfiguration umfaßt hier also auch alle Transienten, die in diesem Bassin verlaufen.

Zu einer anderen Eingangssituation oder zu einer anderen inneren Systemdisposition wird sich gegebenenfalls eine andere semantische Konfiguration einstellen. Das Auftreten einer semantischen Konfiguration kann das Erscheinen weiterer semantischer Konfigurationen in anderen, räumlich getrennten Bereichen zur Folge haben und so zu einer umfassenderen semantischen Konfiguration expandieren. Es kann aber auch das "Verlöschen" anderer Konfigurationen durch Dekohärenzierung auslösen bzw. das Erscheinen einer Kette, d.h. zeitlichen Folge von semantischen Konfigurationen bewirken. Wegen der rekursiven Struktur des Systems kann eine semantische Konfiguration über die ausgelösten Prozesse, die auf sie zurückführen, ihr eigenes Verlöschen zum Ergebnis haben. Diese dynamische Ausdifferenzierung von semantischen Konfigurationen in einem modularen System ist Folge der ausdifferenzierten Verknüpfungsstruktur der Module, die es ihnen erlaubt kooperativ oder auch konkurrierend aufeinander zu wirken. Anknüpfend an die etablierte Begriffsbildung sei angemerkt, daß semantischen Konfigurationen in Form von "neuronalen Ensembles" sichtbar werden, d.h. in Form kohärent feuernder Neuronen der beteiligten Module.

Wie aber entstehen semantischen Konfigurationen in autotropen Systemen? Abgesehen von phylogenetisch vorgegebenen Reiz-Reaktions-Mechanismen, muß das autotrope System adäquates Verhalten in einer sensomotorischen Schleife erlernen. Ob sich eine zu einem gegebenen Muster von Eingangssignalen einstellende kohärente Dynamik von Modulen als verhaltensrelevant, also semantisch erweist, kann erst nach Registrierung ihrer Wirkung auf den motorischen Bereich erkennbar sein. Der Selbstorganisationsprozeß des autotropen Systems muß also, eingebunden in die

sensomotorische Schleife, die inneren Parameter der Module sowie die Kopplungsstärken zwischen den Modulen solange variieren bis sich eine kohärente Dynamik zwischen Modulen einstellt, die eine der äußeren Situation angemessene, von den Effektoren zu erbringende Verhaltensleistung initiiert. An der Ausprägung von semantischen Konfigurationen werden also nicht nur die sensorseitigen Signale sondern ebensosehr die durch Rückkopplung auf das kognitive System einwirkenden motorseitigen Signale sowie innere Mechanismen wie "Aufmerksamkeit", "Bewertung" etc. ihren Anteil haben. Sehr allgemein formuliert, muß Selbstorganisation eine Kohärenz von externer Umweltdynamik und innerer Systemdynamik erzielen. Von welcher Art die biologischen Mechanismen sind, die eine gezielte Veränderung der Systemdynamik bewirken, ist zur Zeit nicht im Detail bekannt, auch wenn solche Lernprozesse auf der Grundlage synaptischer Plastizität durchaus zu verstehen sind (Singer 1991).

Wir haben bislang ein entwickeltes autotropes System betrachtet, also eines, dessen Struktur bereits gegeben ist. Ein weiterer Gesichtspunkt betrifft die untere Zeitskala der Entwicklung eines solchen Systems, die der Adaptation. Hier findet ein Prozeß statt, in dem Neuronen hinzugefügt oder entfernt, neue Verbindungen aufgebaut und bestehende abgebaut werden. Bei diesem Prozeß können schon kleine lokale Änderungen in der (Modul-) Architektur zu drastischen Änderungen des globalen Systemverhaltens führen. Wie immer der Selbstorganisationsprozeß auf dieser Ebene formal zu beschreiben ist, er wird den folgenden Bedingungen genügen müssen (vgl. z.B. Quartz und Sejnowski 1996). Erstens: Das Einbinden oder Ausgliedern von Strukturelementen muß in dem entsprechenden lokalen Maßstab so erfolgen, daß die dynamischen (wir greifen vor) "Repräsentationseigenschaften" des Gesamtsystems sich nicht grundsätzlich ändern (*Lokalitätsbedingung*). Zweitens: lokale Änderungen dürfen unter normalen Umständen zuvor Gelerntes nicht löschen (*Stabilitätsbedingung*). Als Ziel dieser Adaptation wäre die Vergrößerung der "repräsentationalen" Kapazitäten des Systems zu postulieren, was z.B. die Ausdifferenzierung der Modulfunktionen, die Erzeugung von mehr (koexistenten) Attraktoren bzw. die Bereitstellung von mehr semantische Konfigurationen bedeuten kann.

Die Dynamik gekoppelter Module ist der mathematischen Analyse zur Zeit noch schwer zugänglich und nur in speziellen Fällen durchführbar. So z.B. bei der Kaskadierung konvergenter Netzwerke (Hirsch 1989) oder beim Vorliegen definierter Symmetrien in den Kopplungsstrukturen (Atija und Baldi 1989; Collins und Stewart 1994). Ansätze wie z.B. der "coupled map lattice approach" (vgl. Kaneko 1994) zeigen jedoch, daß gekoppelte Systeme raum-zeitlichen Eigenschaften aufweisen, die für die kognitive Neurodynamik vermutlich relevant sind; z.B. die Synchronisation von Teilsystemen, oder die Bildung von "Neuronenclustern" nach feststehenden Phasen, Amplituden und/oder Frequenzbeziehungen. Das Auffinden eines möglichen Zusammenhangs zwischen geeigneten Verknüpfungsstrukturen und relevanten raum-zeitlichen dynamischen Effekten wird für eine Weile noch das Feld von Comptersimulationen geschickt gewählter, biologisch inspirierter Modelle bleiben.

6 Innere Repräsentation

Folgt man dem skizzierten dynamischen Ansatz zur Beschreibung von Gehirnen, so stellt sich im Zusammenhang mit kognitiven Prozessen die Frage nach den inneren Repräsentationen von Strukturelementen der äußeren Welt vollkommen neu. Wie oben erwähnt, wird der Begriff der inneren Repräsentation oft mit der Computer-Metapher des Gehirns verbunden, d.h. mit der Vorstellung, das Gehirn sei ein informationsverarbeitendes System. Die Repräsentation wird dabei als durch (physikalische) Symbole realisiert gedacht. Der Konnektionismus, der sich primär mit konvergenten Netzwerken auseinandersetzte, ging einen Schritt darüber hinaus. Hier wurden Klassen von Signalkonstellationen durch die Gewichte eines Netzwerkes repräsentiert; man sprach von einer sogenannten verteilten oder sub-symbolischen Repräsentation. Beiden Ansichten ist gemeinsam, daß Repräsentation als etwas Statisches begriffen wird. Diese "klassischen" Konzeptionen denken Repräsentation als Re-Präsentation (Vergegenwärtigung) einer Form in einer anderen und verweisen damit implizit auch auf das materielle Substrat des Repräsentierenden; z.B. auf die binäre Repräsentation in digitalen Computern in Form von 0/1-Zuständen ihrer Transistoren (die bisweilen gleichgesetzt wurden mit dem "Feuern" und "Nichtfeuern" von Neuronen in biologischen Gehirnen). So gedacht wird Repräsentation statisch und speicherbar und erlaubt die Vorstellung bzw. Realisierung von Gedächtnis als "Informationsspeicher" z.B. in der Form heute verwendeter Aufzeichnungsmedien.

Innere Repräsentation in biologischen Gehirnen zielt aber eher auf die Re-Generierung Bedeutung tragender neuronaler Signalfolgen als auf das "Abbild" schon existenter "objektiver" Gegebenheiten der externen Welt. Bedeutung von Signalfolgen kann ein situiertes System, also eines, das im Rahmen einer sensomotorischen Schleife agiert, nur für und aus sich selbst heraus schaffen, gegebenenfalls in sozialer Interaktion mit anderen Individuen seiner Population (Roth 1992). Dies ist der wesentliche Beitrag, den der reflexive, situierte Selbstorganisationsprozeß zu leisten hat, und dies ist auch der Kern jenes Prozesses, den man Adaptation oder in spezifischerer Form und auf einer anderen Zeitskala, Lernen nennt. Setzt man ferner voraus, daß in Gehirnen (zumindest in ihrem kognitiven Teilbereich) stationäre Systemzustände nie eingenommen werden, selbst beim Vorliegen stationärer Muster von Eingangssignalen, so kann das klassische Konzept einer statischen Repräsentation kaum aufrecht erhalten werden. Vielmehr haben wir die "repräsentationalen" Eigenschaften eines kognitiven Systems dann als Ergebnis der dynamischen Interaktion zwischen einer strukturierten Umwelt und dem Selbstorganisationsprozeß eines autotropen Systems zu verstehen. Damit beziehen sich innere Repräsentationen aber sowohl auf Strukturelemente der Umwelt, und damit auf die jeweiligen Problemfelder mit denen das System konfrontiert wird, als auch auf die physischen Eigenschaften des Lebewesens selbst, d.h. auf die materielle Beschaffenheit und Struktur seiner Sinnesorgane, seines Bewegungsapparates und seines kognitiven

Systems. Bei der Herausbildung von inneren Repräsentationen werden außerdem systemintrinsische "Bedürfnisse" wie die nach Energieaufnahme oder sexueller Reproduktion eine Rolle spielen. Im Rahmen eines neurodynamischen Verständnisses ist Repräsentation dann nicht mehr als etwas Statisches zu verstehen, das in irgendeinem Medium, und sei es die biochemische "Wetware", räumlich gespeichert ist, also auch nicht als etwas Subsymbolisches, das in Form von Verknüpfungsstrukturen vorliegt. Repräsentation ist vielmehr als etwas Flüchtiges zu sehen, als etwas, das erscheint und wieder vergeht in der Form dynamischer Muster von Hirnaktivität. Sie ist Ausdruck einer Kohärenz zwischen internen neurodynamischen Prozessen und der Dynamik der äußeren Umwelt, die sich in den Veränderungen der Rezeptoraktivitäten widerspiegelt. In diesem Sinne schaffen Repräsentationen verhaltensrelevante Relationen zwischen externen und internen Dynamiken, d.h. sie sind gerichtet auf und bestimmt durch ein angepaßtes, überlebensicherndes Verhalten. In dem Maße wie sich die Strukturelemente der äußeren Umwelt ändern und damit adäquates Verhalten variiert, ändern sich auch die Randbedingungen möglicher Relationen und die damit verbundenen Repräsentationen. Repräsentiert wird auf kognitiver Ebene also nicht "die Welt", sondern jener Teil des auf die Sensorik treffenden Stroms von Photonen, Molekülen, Schallwellen etc., der sich im Rahmen eines Adaptations- bzw. Lernprozesses für ein Überleben in der vorgefundenen ökologischen Nische als relevant erwiesen hat.

Der klassische Begriff der Repräsentation wird aber noch aus einem zweiten Grund obsolet: Er verweist als Abbildung von einer materiellen Form in eine andere immer auch auf die Strukturelemente des repräsentierenden Substrats. Solche Strukturelemente sind, als "Ziel" dieser Abbildungen auf das entsprechende Substrat (man denke z.B. an die photosensiblen Emulsionen der Fotografie oder die Transistorarrays der Computer), immer schon Teil der Repräsentation. Versteht man in bezug auf Gehirne unter einem Strukturelement eine Gruppe von "verschalteten" Neuronen, z.B. im Sinne von Modulen oder Modulaggregaten, so ergibt sich die folgende Schwierigkeit: Aufgrund der Multifunktionalität dieser Strukturelemente kann ein und dasselbe Element bei verschiedenen sensorischen Eingangsmustern in den resultierenden globalen Prozessen unterschiedliche Funktionen erfüllen, d.h. an der Repräsentation verschiedenster Aspekte beteiligt sein. Weiters kann ein und dieselbe Funktion von sehr unterschiedlichen Strukturelementen erfüllt werden. Damit verliert aber Repräsentation das Charakteristikum, an spezifische statische Strukturelemente gebunden zu sein und eine strukturelle systemintrinsische Bedeutung zu besitzen.

In bezug auf die Beschreibung kognitiver Systeme als autotrope Systeme werden wir im folgenden den Begriff der inneren Repräsentation im Sinne einer *semantischen Konfiguration* von Moduldynamiken verwenden. Als solche gewinnt, oder besser, erzeugt sie "Bedeutung" nur in jenem flüchtigen Moment, in dem sie effektiv, d.h. zur Vollbringung einer verhaltensrelevanten Leistung, genutzt werden kann. Als semantische Konfiguration ist eine innere Repräsentation nirgends gespeichert, sie wird erst durch die und in der globalen Dynamik eines kognitiven Prozesses entfaltet. Sie

kann bestenfalls als stabile dynamische Teilstruktur einer kognitiven Dynamik verstanden werden, die vermutlich nur in Zeitintervallen von einem Sekundenbruchteil bis zu wenigen Sekunden existiert. Sie ist damit nicht nur nicht-statisch sondern auch als Dynamik nicht-persistent. Ferner ist sie im Unterschied zur klassischen Konzeption von Repräsentation nicht "passiv", d.h. nur Abbild von Strukturelementen der Umwelt. Als Teildynamik eines kognitiven Prozesses besitzt sie eine aktive Funktion, die darin besteht, Überleben sicherndes Verhalten zu generieren. Als kognitiver Teilprozeß wird sie von Prozessen hervorgerufen und bewirkt eine Folge von Prozessen.

Wenn eine Repräsentation als semantischen Konfiguration in einem kognitiven Prozeß aufscheint, dann ist davon auszugehen, daß sie nie als distinkte, immer gleiche Teildynamik zu identifizieren ist. Sie wird als Transientendynamik in der Form eines immer wieder anderen dynamischen Prozesses ihre Funktion erfüllen. Wir haben angenommen, daß letztlich das Bassin des jeweiligen Attraktors entscheidend für ihre Wirkung ist. Damit sind aber auch die durch eine semantische Konfiguration hervorgerufenen Effekte unter verschiedenen inneren und äußeren Bedingungen, d.h. unter den verschiedenen, durch das Bassin definierten Anfangsbedingungen, reproduzierbar. Die Attraktoren selbst, und damit auch ihre Bassins und die zugehörigen semantischen Konfigurationen, sind aber veränderlich. Sie hängen davon ab, in welchem Parameterbereich die beteiligten Module operieren. Eine spezifische semantische Konfiguration kann also nur dann in Erscheinung treten, wenn erfahrungs-, situations- und bedürfnisbedingt eine entsprechende Konstellation von Signalen an den Parametereingängen (wieder) anliegt. Dies bedeutet, daß die innere Repräsentation von Strukturelementen der äußeren Welt vom momentanen Zustand anderer Systemgruppen abhängen kann. Ferner folgt aus dem Handlungsbezug von inneren Repräsentationen, daß ein Gehirn bei der Wahrnehmung einer sensorseitigen Signalkonstellation eine semantische Konfiguration wählen muß, in der die Erzeugung von motorischen Signalen zu angemessenem Verhalten einbegriffen ist. Bei der Konstituierung geeigneter Konfigurationen wird also auch das "Monitoring" körpereigener Positionen und Bewegung einen Anteil haben. So muß z.B. die reale physikalische Bewegung von externen Objekten von der durch die Motorik erzeugten "virtuellen" Bewegung auf der peripheren Sensorfläche unterschieden werden. Ein Hinweis auf das enge Zusammenspiel von sensorischen und motorischen Einflüssen bei der Konstituierung innerer Repräsentationen ist die Tatsache, daß in Gehirnen Areale gefunden werden, die Signale sowohl von der sensorischen als auch von der motorischen Seite her erhalten. Innere Repräsentation erscheint somit als Vergegenwärtigung schon dagewesener Konfigurationen von Eingangssignalen in Verbindung mit Konstellationen von Ausgangssignalen, die antrainierten oder positiv bewerteten Verhaltensmöglichkeiten entsprechen. An der Konstituierung von innerer Repräsentation als semantische Konfiguration sind also nicht nur sensorseitige Signale, sondern ebenso solche des motorischen Bereichs sowie solche anderer, innerer Teilsysteme (Aufmerksamkeit, Bewertung, Bedürfnisse usw.) beteiligt. Innere Repräsentation ist daher nicht mehr als purer "Reflex", als Abbild einer "objektiv" gegebenen physikalischen Umwelt zu verstehen.

Das Auftreten von semantischen Konfigurationen im Rahmen eines kognitiven Prozesses kann einerseits durch die von der Umwelt verursachten sensorseitigen Signale und andererseits, und vermutlich in viel stärkerem Maße, durch die "restliche" innere Dynamik des gesamten Systems bedingt sein. Durch den Einfluß der Aktivität anderer Systemkomponenten können über semantische Konfigurationen die gleichen Effekte hervorrufen werden wie durch das Vorliegen einer entsprechenden äußeren Stimulation. Als situierte adaptive, d.h. überlebensfähige Systeme werden kognitiver Systeme durch diese *autonomen* inneren Prozesse in die Lage versetzt, Vorhersagen zu machen und sinnvolle Handlungsstrategien zu entwickeln, d.h. *prädiktive Weltmodelle* zu erstellen. Innere Repräsentationen als intern erzeugte Konfigurationen von kohärenten Moduldynamiken sind dann als Bausteine eines Weltmodells zu verstehen, auf dessen Grundlage eine innere Exploration von Handlungsalternativen erfolgen kann. Demnach entspricht einer jeden solchen Konfiguration einer Menge von Aspekten der Umwelt, so wie sie von den Sensoren gefaßt und von der Motorik "manipuliert" werden können. Als Teildynamiken eines kognitiven Prozesses sind sie immer wieder neu assemblierbar, und um konsistente Weltmodelle zu ergeben müssen sie miteinander "verträglich" sein. In diesem Sinne sind semantische Konfigurationen als Hypothesen zu fassen, die getestet, bestätigt oder verworfen werden müssen. Als modulare neurodynamische Prozesse werden sie also miteinander konkurrieren. Ein Kriterium für die Güte einer semantischen Konfiguration als Hypothese ist ihre Brauchbarkeit für das Lebewesen in der *Zukunft*. Erfolgreiche Konfigurationen repräsentieren in diesem Sinne Gesetzmäßigkeiten der externen dynamischen Abläufe, oder anders gesagt, sie sind zugleich kohärent, d.h. in Einklang mit der externen Dynamik.

Diese Fähigkeit kognitiver Systeme zu autonomen internen Prozessen verleitet vermutlich zu der radikalen systemtheoretischen Annahme, Gehirne könnten als (irgendwie) abgeschlossene Systeme beschrieben werden, deren einziges Anliegen darin besteht, ihren eigenen Zustand zu optimieren. Der Einfluß der äußeren Welt auf das kognitive System wird so auf den einer reinen Störungen der internen Dynamik reduziert, d.h. auf ein "externes Rauschsignal", welches das System permanent zwingt über sein motorisches Instrumentarium den angestrebten Zustand wiederherzustellen. Ein solcher Ansatz entbindet den Begriff der inneren Repräsentation jeder beschreibungsrelevanten bzw. erkenntnistheoretischen Bedeutung. Dies auch dann, wenn die Umwelt letztlich doch, auf einer anderen Zeitskala, im Rahmen der Adaptation Einfluß auf die materielle Strukturentwicklung des Systems nimmt.

Wir können unsere Überlegungen wie folgt zusammenfassen: Innere Repräsentationen biologischer Gehirne sind als Teilprozesse eines globalen kognitiven Prozesses gegeben und als solche nicht statisch und nicht persistent. Sie existieren einzig in kleinen Zeitfenstern als reproduzierbare semantische Konfigurationen von Moduldynamiken. In dieser Form sind sie nicht speicherbar und nur indirekt und nicht eindeutig durch Verknüpfungsstrukturen des Gehirns bestimmt. Sie stellen keine Widerspiegelungen einer "objektiven" externen Welt dar, sondern sind bedingt durch die sensomotorische

Ausstattung eines Lebewesens und (primär) bezogen auf die selbstgenerierte Bedeutung überlebenswichtiger Aspekte der jeweils vorgefundenen Umwelt. In Abwandlung des klassischen Konzepts von innerer Repräsentation, könnte man sie als Re-Generierung einer Abbildung von einer zeitlichen Form (externe Dynamik) in eine andere zeitliche Form (Kohärenz von Moduldynamiken) charakterisieren.

7 Zusammenfassung

Die Ausführungen zu inneren Repräsentationen im Rahmen des dynamischen Ansatzes zur Beschreibung kognitiver Hirnprozesse müssen zur Zeit weitgehend auf der deskriptiven Ebene erfolgen, da eine konsistente, mathematisch fundierte "Theorie des Gehirns" nicht ausgearbeitet ist. Ihre Formulierung wird große Anstrengungen erfordern, sowohl was die Entwicklung ihres begrifflichen Teils als auch ihres mathematischen Instrumentariums anbelangt. Dieser Beitrag sollte als Versuch gewertet werden, zu erkunden, inwieweit die Vorstellung einer inneren Repräsentation unter dynamischen Gesichtspunkten noch tragfähig ist, wenn man denn diesen Begriff im Kontext der Hirnforschung nicht von vornherein als unzulänglich bezeichnet und darum vermeiden will. Denn die Vorstellung von der Existenz innerer Repräsentationen der externen Welt beruht letztlich auf unserer menschlichen Introspektion und Sprachfähigkeit. Es ist jedoch keineswegs sicher, daß die tatsächlich ablaufenden Prozesse eine dieser Vorstellung gemäße Entsprechung haben. Unsere Konkretisierung von innerer Repräsentation als semantische Kohärenz von Moduldynamiken in einem autotropen System wird aber nur ein sprachlicher Kunstgriff und eine neurowissenschaftliche Schimäre bleiben, wenn deren Entstehungs-, Existenz- und Wirkungsbedingungen auf materieller und formaler Ebene nicht verstanden und nachvollzogen werden können. Insofern stehen innere Repräsentationen ebenso wie ihre phänomenologischen Erscheinungsformen als "neuronale Ensembles" auch in Zukunft im Zentrum neurowissenschaftlicher Fragestellungen. Entscheidend für ein tieferes Verständnis des dynamischen Ansatzes wird sein, jene Regeln zu finden, nach denen sich geeignete innere Repräsentationen als kohärente Moduldynamiken konstituieren und so "zueinanderfinden", daß sie ein überlebenssicherndes Verhalten generieren können.

Grundlage unserer Überlegungen war die Beschreibung von Gehirnen als autotrope Systeme, die wir als modulare Systeme vorausgesetzt haben. Die Einführung chaotischer Neuromodule als elementare Systemelemente war motiviert durch das Anliegen, Multifunktionalität bzw. kontextabhängige Verhaltensvariabilität schon auf der untersten Beschreibungsebene kognitiver Systeme zu ermöglichen. Wir haben versucht deutlich zu machen, daß komplexe dynamische Eigenschaften schon mittels relativ einfacher neuronaler Strukturen realisiert werden können, die aus exzitatorischen und inhibitorischen Neuronen bestehen. In bezug auf Gehirne als autotrope Systeme wäre - salopp und computermetaphorisch formuliert - die Schlußfolgerung zu ziehen,

daß nicht nur die "Wetware" (Neuronen und ihre Verknüpfungsstruktur) komplex ist, sondern daß die verwendete "Software" (die neurodynamischen Prozesse) ungleich komplexer ist. Dies ist vermutlich die Barriere, an der ein monokausaler, rein "mechanistischer" oder auch informationstheoretischer Zugang zur Beschreibung des Gehirns seine Grenzen findet.

Literatur

Abeles, M. (1991), *Corticonics: Neural Circuits of the Cerebral Cortex*, Cambridge: Cambridge University Press.

Abraham, R.H. and Shaw, C.D. (1992), *Dynamics - The Theory of Behavior*, 2nd. ed., Redwood City: Addison-Wesley.

Aihara, K., Takabe, T. and Toyoda, M. (1990), Chaotic neural networks, *Physics Letters A*, **144**,333-340.

an der Heiden, U. (1980), *Analysis of Neural Networks*, Lecture Notes in Biomathematics **35**, Berlin: Springer.

Atiya, A. and Baldi, P. (1989), Oscillations and synchronizations in neural networks: An exploration of the labeling hypothesis, *International Journal of Neural Systems*, **1**,103-124.

Babloyantz, A. and Destexhe, A. (1986), Low-dimensional chaos in an instance of epilepsy, *Proc. Nat. Acad. Sci. USA*, **83**, 3513-3517.

Babloyantz, A. and Lourenço, C. (1994), Computation with chaos: A paradigm for cortical activity, *Proc. Natl. Acad Sci. USA*, **91**, 9027-9031.

Baird, B. (1990), Bifurcation and category learning in network models of oscillating cortex, *Physica*, **D42**, 365-384.

Brooks, R.A. (1991), Intelligence without representation, *Artificial Intelligence*, **47**, 139-159.

Buzsáki, G., Llinás, R., Singer, W., Berthoz, A. and Christen, Y. (eds.) (1994), *Temporal Coding in the Brain*, Berlin: Springer.

Collins, J.J. and Stewart, I. (1994), A group-theoretic approach to rings of coupled biological oscillators, *Biological Cybernetics*, **71**, 95-103.

Duke, W. and Pritchard, W.S. (eds.) (1991), *Proceedings of the Conference on Measuring Chaos in the Human Brain*, Singapore: World Scientific.

Elbert, T., Ray, W.J., Kowalik, Z.J., Skinner, J.E., Graf, K.E. and Birbaumer, N. (1994), Chaos and physiology: Deterministic chaos in excitable cell assemblies, *Physiological Review,* **74**, 1-47.

Ermentrout, B. (1984), Period doublings and possible chaos in neural models, *SIAM J. Appl. Math.*, **44**, 80-95.

FitzHugh, R. (1969), Mathematical models of excitation and propagation in nerve, in: Schwan, H.P. (ed.), *Biological Engineering*, New York: McGraw-Hill.

Freeman, W.J. (1975), *Mass Action in the Nervous System*, New York: Academic Press.

Freeman, W.J. (1992), Tutorial on neurobiology: From single neurons to brain chaos, *International Journal of Bifurcation and Chaos*, **2**, 451-482.

von Foerster, H. and Zopf, G.W. (eds.) (1962), *Prinziples of Self-Organization: The Illinois Symposium on Theory and Technology of Self-Organizing Systems*, London: Pergamon Press.

Gerstner, W. , Ritz, R. and van Hemmen, J.L. (1993), Why spikes? Hebbian learning and the retrieval of time-resolved excitation patterns, *Biol. Cybern.*, **69**, 503-515.

Guevara, M.R., Glass, L., Mackey, M.C. and Shrier, A. (1983), Chaos in neurobiology, *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC13**, 790-798.

Glünder, H. (1993), Zentralnervöse Repräsentationen bei der sensomotorischen Informationsver- arbeitung: ein Plädoyer für verhaltens-relevante Konzepte, *Kognitionswissenschaft*, **3**, 127-138.

Haken, H. (1982), *Synergetics*, Berlin: Springer.

Haken, H. (1995), *Principles of Brain Functioning*, Berlin: Springer.

Harth, E., Csermely, T. J., Beek, B. and Lindsay, R. D. (1970), Brain functions and neural dynamics, *J. Theoret. Biol.*, **26**, 93-120.

Harth, E. (1983), Order and chaos in neural systems: An approach to the dynamics of higher brain funktions, *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC13,** 782-789.

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

Hertz, J., Krogh, A. and Palmer, R.G. (1991), *Introduction to the Theory of Neural Computation*, Redwood City: Addison-Wesley.

Hirsch, M. W. (1989), Convergent activation dynamics in continuous time networks, *Neural Networks*, **2**, 331-350.

Hodgkin, A.L. and Huxley, A.F. (1952), A quantitative description of membrane current and its application to conduction and excitation in nerve, *J. Physiol.*, **117**, 500-544.

Holland, J.H. (1995), *Hidden Order - How Adaptation Builds Complexity*, Reading: Addison-Wesley.

Jackson, E.A. (1991), *Perspectives of Nonlinear Dynamics*, Vol. I, Cambridge: Cambridge University Press.

Kaneko, K. (1994), Relevance of dynamical clustering to biological networks, *Physica*, **D75**, 55-73.

Kelso, J. A. S. (1995), *Dynamic Patterns - Self-Organization of Brain and Behavior*, Cambridge: MIT Press.

Maturana, H. and Varela, F.J. (1987), *Der Baum der Erkenntnis. Die Biologischen Wurzeln des Erkennens*, München: Scherz.

Mazoyer, B., Roland, P. and Seitz, R. (eds.) (1995), International Congress on the Functional Mapping of the Human Brain, *Human Brain Mapping*, **Suppl. 1**.

Minski, M.L. and Papert, S.A. (1969), *Perceptrons*, Cambridge: MIT Press, expanded edition: 1988.

Nagumo, J, Arimoto, S. and Yoshizawa, S. (1962), An active pulse transmission line simulating nerve axon, *Proc. IRE*, **50**, 2061-2070.

Ott, E. (1993), *Chaos in Dynamical Systems*, Cambridge: Cambridge University Press.

Ott, E., Grebogi, C. and Yorke, J.A. (1990), Controlling Chaos, *Phys. Rev. Lett.* **64**, 1196-1199.

Pantev, C., Elbert, T. and Lutkenhöhner B. (eds.) (1995), *Oscillatory event-related brain dynamics*, London: Plenum.

Port, R.F. and van Gelder, T. (1995), *Mind as Motion - Explorations in the Dynamics of Cognition*, Cambridge: MIT Press.

Quartz, S.R. and Sejnowski, T.J. (1996), The neural basis of cognitive development: A constructivist manifesto, *Behavioral & Brain Sciences*, to appear.

Reeke, G.N. and Sporns, O. (1993), Behaviorally based modelling and computational approaches to neuroscience, *Annual Review of Neuroscience*, **16**, 597-623.

Roth, G. (1992), Kognition: Die Entstehung von Bedeutung im Gehirn, in: Krohn, W. und Küppers, G., *Emergenz: Die Entstehung von Ordnung, Organisation und Bedeutung*, Frankfurt: Suhrkamp.

Roth, G. (1994), *Das Gehirn und seine Wirklichkeit*, Frankfurt: Suhrkamp.

Rumelhart, D.E. and McClelland, J.L. (eds.) (1986), *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, Vol. 1,2, Cambridge: MIT Press.

Shimizu, H. (1993), Biological autonomy: the self-creation of constraints, *Applied Mathematics and Computation*, **56**, 177-201.

Skarda, C.A. and Freeman, W.J. (1987), How brains make chaos in order to make sense of the world, *Behav. Brain Sci.* **10**, 161-195.

Singer, W., (1991), Die Entwicklung kognitiver Strukturen - ein selbstreferenzieller Lernprozeß, in: Schmidt, S.J. (Hrsg.), *Gedächtnis*, Frankfurt: Suhrkamp.

Singer, W., (1993), Synchronization of cortical activity and its putative role in information processing and learning, *Annual Review of Physiology*, **55**, 349-374.

Singer, W., (1995), Time as coding space in Neocortical Processing: A hypothesis, in: M.S. Gazzaniga (ed.), *The Cognitive Neurosciences*, Cambridge: MIT Press.

Stadler, M. und Kruse, P. (1992), Zur Emergenz psychischer Qualitäten, in: Krohn, W. und Küppers, G. (Hrsg.), Die Entstehung von Ordnung, Organisation und Bedeutung, Frankfurt: Suhrkamp.

Steels, L. and Brooks, R. (eds.) (1995), *The Artificial Life Route to Artificial Intelligence: Building Embodied, Situated Agents*, Hillsdale: Lawrence Erlbaum.

Stollenwerk, N. and Pasemann, F. (1996), Control strategies for chaotic neuromodules, *International Journal of Bifurcation and Chaos*, **6**, 693 - 702.

Szentágothai, J. (1983), The modular architectonic principle of neural centers, *Rev. Physiol. Biochem. Pharmacol.*, **98**, 11-61.

Thatcher, R.W., Hallett, M., Zeffiro, T., John, E.R. and Huerta, M. (eds.) (1994), *Functional Neuroimaging - Technical Foundations*, San Diego: Academic Press.

Tsuda, I. (1991), Dynamik link of memory - Chaotic memory map in nonequilibrium neural networks, *Neural Networks*, **5** , 313 - 326.

Varela, F.J., Thomson E. and Rosch, E. (1992), *Der Mittlere Weg der Erkenntnis*, München: Scherz.

Varela, F.J. (1994), *Ethisches Können*, Frankfurt: Campus.

von der Malsburg, C. (1981), The correlation theory of brain function, *Internal Report* **81**-2, Abteilung für Neurobiologie, MPI für Biophysikalische Chemie, Göttingen.

Wennekers, T. and Pasemann F. (1996), Synchronous chaos in highdimensional modular neural networks, *International Journal of Bifurcation and Chaos*, **6**, 2055 - 2067.

Wiggins, S. (1990), *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, Texts in Applied Mathematics 2, New York: Springer.

Wilson, H.R. and Cowan, J.D. (1972), Excitatory and inhibitory interactions in localized populations of model neurons, *Biophysical Journal*, **12**, 1-24.

Yovits, M.C. and Cameron, S. (eds.) (1960), *Self-Organizing Systems*, London: Pergamon Press.