

# Autonomie in Biologie und Technik

## Kognitive Netzwerke - Artificial Life - Robotik

E. von Goldammer & J. Paul [\*]

### I. Ausgangs-Situation

Der Versuch, eine wissenschaftliche Beschreibung lebender Systeme im Sinne einer "ganzheitlichen", d.h. nicht-reduktionistischen 'Theorie des Lebens' zu entwickeln, hat in den 70er Jahren zu einem fundamentalen Wechsel des bis dahin gültigen wissenschaftlichen Paradigmas einer strikten Trennung von Beobachter und Beobachtetem geführt. Der Beobachter wird zum "Teil des zu beschreibenden Systems" <sup>1</sup>:

"Ein lebender Organismus ist eine selbständige autonome, organisatorisch geschlossene Wesenheit,  
UND  
ein lebender Organismus ist selbst Teil, Teilhaber und Teilnehmer seiner Beobachtungswelt." (1)

Diese beiden sich zueinander komplementär verhaltenden Aussagen setzen zunächst einmal 'Autonomie', d.h. 'Selbst-Regelung' für lebende Systeme notwendig voraus. Dabei ist der Begriff der 'Selbst-Regelung' synonym mit dem Ausdruck 'Regelung der Regelung', und das bedeutet in der Terminologie der Kybernetik:

"Ein lebendes System regelt seine Regelung (selbst)." (2)

Die Akzeptanz einer derartigen Aussage hat erhebliche Konsequenzen für die (kybernetische) Beschreibung autonomer Systeme, denn sie verlangt die organisatorische Geschlossenheit autonomer Systeme im Sinne der "Closure Thesis" <sup>2</sup>:

Closure Thesis:  
"Every autonomous system is organizationally closed.  
... organizational closure is to describe a system with no input and no output ..." (3)

Diese Anschauung ist mit dem *Wiener'schen* Begriff des "Feedback" schlechthin unvereinbar. Hier wird der Übergang von der klassischen Kybernetik (1. Ordnung), deren Beschreibungsobjekte ausschließlich 'Input/Output'-Systeme sind, zur "Kybernetik 2. Ordnung" deutlich, die es ganz offensichtlich mit (operativ) geschlossenen, d.h. autonomen Systemen zu tun hat.

Der epistemologisch entscheidende Punkt resultiert aus der Erkenntnis, dass 'operative Geschlossenheit' und 'Autonomie' lebender Systeme unvereinbar sind mit einer Beschreibung des Systems aus seinem (vom Beobachter festgelegten) System/Umgebungs-Verhältnis heraus. Mit anderen Worten, die durch den Beobachter definierte Abgrenzung von System

---

\* publiziert in: Selbstorganisation – Jahrbuch für Komplexität in Natur-, Sozial- und Geisteswissenschaften, Band 6: "Realitäten und Rationalität" (A.Ziemke und R.Kaehr, hrsg.), Duncker & Humblot, Berlin 1995, p.277-298.

<sup>1</sup> von Foerster, H., Kybernetik einer Erkenntnistheorie, in: Sicht und Einsicht, Vieweg Verlag, Braunschweig, 1985.

<sup>2</sup> Varela, F.: Principles of Biological Autonomy, in: General Systems Research (G. Klir, ed.), Vol. II, North Holland Publ., Amsterdam, 1979.

und Umgebung, durch die ein Input/Output-Verhältnis erst definiert wird, ist immer unterschiedlich zur Grenzbildung, die das autonome System durch seine (operative) Geschlossenheit relativ zu allen anderen Systemen erzeugt. Es existieren also zwei vollständig unterschiedliche, nicht zu vereinbarende (standortabhängige) Beschreibungsvarianten:

- einmal vom Standpunkt eines externen Beobachters eines (von ihm definierten) Systems und dessen Umgebung aus,  
 UND (4)  
 vom Ort des autonomen (lebenden) Systems selbst (unter Einbeziehung des Beobachters) heraus, also der vom autonomen System festgelegten Grenz-  
 ziehung zwischen sich und seiner Umgebung.

Parallel zur Entstehung der besonderen Rolle des Beobachters in der Konzeption einer 'Theorie lebender Systeme' steht in diesem Kontext die Fragestellung nach der Relation von System und Umgebung, die wiederum unter dem Aspekt der kognitiven Fähigkeiten als primordiale Eigenschaft von 'Leben überhaupt' verstanden wird<sup>3</sup>:

- "Lebende Systeme sind kognitive Systeme, und Leben als Prozess ist ein Prozess der Kognition. Diese Aussage gilt für alle Organismen, ob diese ein Nervensystem besitzen oder nicht." (5)

Auf dem Weg zu einer 'Theorie lebender Systeme' kommt der Konzeption der 'Autopoiese' von *Maturana* und *Varela*<sup>4</sup> eine zentrale Rolle zu. Dabei stellt die 'Theorie autopoietischer Systeme' den Versuch einer rein semantischen, d.h. nicht-formalen Theorie lebender Systeme dar, mit der erklärten Absicht, eine biologische (nicht-physikalistische) Begrifflichkeit lebender Systeme zu entwickeln – das ist ihr Verdienst. Was auf der Basis einer rein semantischen Theorie jedoch nicht gelingen kann, ist eine Symbiose von Computer- und Biowissenschaften im Sinne der Simulation biologischer Systeme und der daraus resultierenden Konstruktion entsprechender technischer Artefakte – das ist das Problem.

## II. Wissenschaftlich-Technische Problemstellung

Während alle bis heute bekannten Modelle der Neuroinformatik ausschließlich klassische Input/Output-Systeme – also offene Systeme – beschreiben, stellen die von der 'Kybernetik 2. Ordnung' geforderten Modelle biologisch kognitiver Netzwerke geschlossene Systeme dar. Hier besteht ganz offensichtlich ein unvereinbarer Widerspruch in der Vorstellung zwischen 'offenen' und 'geschlossenen' Systemen, Netzwerken oder Modellen der Beschreibung. Entscheidend ist dabei die Erkenntnis, dass nur geschlossene Systeme eine Umgebung besitzen können, während offene Systeme prinzipiell keine Umgebung besitzen.

Wird also nach einem Modell zur Beschreibung kognitiver Prozesse gesucht, dann muss dieses den Aspekt der 'Geschlossenheit' beinhalten, denn

- Kognition ist die Fähigkeit eines Systems aus eigener Leistung zwischen sich und seiner Umgebung eine Unterscheidung treffen zu können. (6)

<sup>3</sup> Maturana, H.: *Biologie der Kognition*, in: *Erkennen, die Organisation und Verkörperung von Wirklichkeit*, Vieweg Verlag, Braunschweig, 1975.

<sup>4</sup> Maturana, H. & Varela, F.: *Autopoiesis: The Organization of the Living*, in: *Autopoiesis and Cognition*, Boston Studies in Philosophy of Science, Vol. 42, p.63-134, (M.S.Cohen, M.W. Wartofsky, eds.) D. Reidel Publ., Dordrecht 1972.

Dies wiederum setzt notwendigerweise die Existenz einer Umgebung für das System – vom Standpunkt des Systems aus – voraus und nicht lediglich nur vom Standpunkt eines Beobachters des Systems aus, wie dies heute in aller Regel beim Begriff 'Umgebung' naiverweise angenommen wird (siehe dazu Aussage 4). Es sei hier der Hinweis gestattet, dass bei dieser Definition von Kognition zwischen 'Kognition' und 'Bewusstsein' unterschieden wird. Um bei einem System von 'Bewusstsein' zu sprechen, muss dieses über kognitive Fähigkeiten verfügen; die Umkehrung der Aussage, dass kognitive Systeme über 'Bewusstsein' verfügen, ist nicht allgemeingültig. Sich diesen Sachverhalt zu verdeutlichen, ist notwendig, wenn 'Kognition' als eine charakteristische Eigenschaft lebender Systeme angesehen wird (vgl. Aussage 5), durch welche diese sich von den toten Objekten der Physik und Chemie unterscheiden.

Eine Definition der jeweiligen Begriffe wie 'Umgebung' oder 'Kognition' ist, wenn (Kognitions-)Wissenschaft ernsthaft betrieben werden soll, zwingend notwendig. Die Bedeutung einer solchen Definition für technische Entwicklungen kognitiver Systeme lässt sich anschaulich an dem Beispiel eines Roboters demonstrieren:

Betrachtet sei zunächst der Roboter in einem Automobilwerk, dessen Aufgabe darin bestehen soll, Schrauben an einer Karosserie zu befestigen – selbstverständlich handelt es sich hierbei nicht um ein kognitives System. Ein unvoreingenommener Beobachter dieses Roboters wird ohne weiteres eine Unterscheidung zwischen diesem Roboter und seiner Umgebung (den Schrauben, der Karosserie, etc.) treffen können. Vom Standpunkt des Roboters aus (gemäß Aussage 4), stellt sich die Situation jedoch völlig anders dar: dieser Roboter besitzt keine Umgebung. Die "Schrauben", "ihr Weg vom Regal zur Karosserie", etc. sind Teile des Robotprogramms; sie sind sozusagen als Objekte vom Konstrukteur einprogrammiert und gehören zum Computerprogramm, das die Bewegungsabläufe steuert. Letzteres ist für die Funktion dieses technischen Systems ebenso bedeutsam wie die für den Beobachter sichtbare Stahlkonstruktion. Auf der anderen Seite sollte ein zur Kognition befähigter Roboter in der Lage sein, zwischen sich und seiner Umgebung eine Unterscheidung treffen zu können, damit er sich – um im Bilde zu bleiben – nicht selbst auseinander schraubt. Bis heute gibt es noch keine mit kognitiven Fähigkeiten ausgestattete Roboter. Daran ändert sich auch nichts durch die Einführung oder Implementierung von neuronalen Netzwerken, von Fuzzy-Reglern oder gar einer Kombination aus beiden, den so genannten Neuro-Fuzzy-Systemen.

## 1. Physikalische Systeme sind "offene" Systeme

Berücksichtigt man, dass es sich bei der Fuzzy Logik um einen Kalkül und bei den neuronalen Netzwerken um Modelle handelt, denen ein Kalkül zugrunde liegt, so wird schon an dieser Stelle deutlich, dass zumindest aus konzeptioneller Sicht der Begriff des 'Neuro-Fuzzy-Systems' einem wissenschaftlichen Eintopf gleichkommt, in dem Unterschiedliches miteinander verkocht werden soll, in der Hoffnung, dass durch die Einführung von Prinzipien wie 'Selbstorganisation' oder 'Unschärfes Schließen', getreu der Homunkulus-Idee, ein 'intelligentes' System entstehen möge. Ohne eine klare Definition dessen, was beschrieben, modelliert oder konstruiert werden soll, stellen solche oder ähnliche Begriffskombinationen jedoch eine konzeptionell-wissenschaftliche Sackgasse dar.

Die oben gegebene Definition von 'Umgebung' und 'Kognition' erfordert, und das kann gar nicht oft genug betont werden, die Annahme 'geschlossener' Systeme. Um die Konsequenz des Postulats der 'Geschlossenheit' lebender Systeme zu verdeutlichen, ist es angebracht, die entsprechende Begriffsbildung in den Naturwissenschaften einmal zu hinterfragen. Aus Physik und Chemie sind wir gewöhnt, uns über den Systembegriff nur wenig oder gar keine

Gedanken zu machen. 'Offene' und 'geschlossene' Systeme werden allenfalls in der Thermodynamik abgehandelt und dort im allgemeinen als geometrische Abgrenzungen verstanden, d.h., die Systeme werden durch einen Raumbereich definiert. Die Unterscheidung zwischen 'geschlossen' und 'offen' bezieht sich in der Physik darauf, ob Materie, die ein solcher Raumbereich enthält, bei Prozessen in dem Raumbereich verbleibt oder nicht. Ein im Rahmen dieser Begriffsbildung als 'geschlossen' festgelegtes physikalisches System lässt im Gegensatz zu einem 'offenen' keine Materie durch seine Begrenzungen strömen. Ist die Begrenzung eines Systems nicht nur für Materieströme, sondern für alle Energieströme undurchlässig, dann wird ein derartig festgelegtes physikalisches System als 'abgeschlossen' oder 'isoliert' bezeichnet. Man erkennt, dass bei dieser Systemdefinition an eine räumliche Vorstellung appelliert wird. Diese auf den ersten Blick vermeintlich anschauliche Systemdefinition ist jedoch sowohl aus mathematischer wie auch aus physikalischer Sicht nicht nur unzweckmäßig, sondern im höchsten Masse wissenschaftlich inkonsequent. Sie stammt aus einer Zeit, in der die Stoffmenge (gemessen in 'mol') als physikalische Größe allgemein noch nicht akzeptiert war und die 'chemische Energie' als Energieform von den Physikern ignoriert wurde.

Nun haben aber physikalische Systeme immer eine gemeinsame Eigenschaft, nämlich Energie in verschiedenen Formen mit anderen (physikalischen) Systemen auszutauschen. Dabei ändert sich der physikalische Zustand des betrachteten Systems von einem, sagen wir Zustand 1 in einen Zustand 2 – oder anders ausgedrückt von einem Anfangszustand zu einem Endzustand. Die Veränderungen der das System beschreibenden physikalischen Variablen ist das, was in der Physik gemessen wird. Ändert sich der Zustand eines Systems nicht, dann kann man auch nichts messen oder anders herum ausgedrückt, tauscht das betrachtete System keine Energie mit einem anderen System aus, dann ändert sich nichts an dem System und es lässt sich infolgedessen auch nichts messen. Das bedeutet aber im vorliegenden Kontext 'offener' und 'geschlossener' Systeme, dass es in der Physik keinen Sinn macht, Systeme zu betrachten, die keine Energie austauschen können – die Physik (und Chemie) kennt nur Systeme, die offen sind, d.h. einen Austausch von Energie mit anderen Systemen zu lassen. Für die formale Darstellung physikalischer Systeme sind Begriffe wie 'offen' und 'geschlossen' völlig überflüssig,<sup>5</sup> was nicht überraschend ist, da sich beide Begriffe wechselseitig bedingen. Nur wenn auch physikalische Systeme, bei denen keine Energie mit anderen Systemen ausgetauscht werden kann – die im sprichwörtlichen Sinne als geschlossen angesehen werden müssten – eine physikalische Bedeutung hätten, müssten die Begriffe von 'Offenheit' und 'Geschlossenheit' im Zusammenhang mit der Systemfestlegung in einer physikalischen Theorie berücksichtigt werden. Solche hypothetischen "geschlossenen" (physikalischen) Systeme befinden sich in einem physikalischen Zustand, in dem sie bis in alle Ewigkeiten verharren, an ihnen lässt sich nichts messen und damit sind sie aus physikalischer Sicht bedeutungslos. Kurz, der Begriff der 'Geschlossenheit' macht in der Physik und Chemie keinen Sinn.

Für die Systemdefinition in der Physik und Chemie werden die verschiedenen Energieformen, die an dem System ausgetauscht werden, als Summe bilanziert. Man erhält auf diese Weise eine Differentialgleichung – die so genannte Gibbs'sche Funktion – die das damit jeweilig definierte physikalische System vollständig beschreibt, d.h. es handelt sich hierbei

---

<sup>5</sup> Falk, G., & Ruppel, W.: Energie und Entropie – Eine Einführung in die Thermodynamik, Springer Verlag, Berlin, 1976.

um die allgemein gültige Definition eines physikalischen Systems ohne geometrische Begrenzung (näheres siehe Ref. 5):

$$dE = \sum_i \xi_i \cdot dX_i \quad (7)$$

In der Gleichung (7) stehen links die Änderung der Gesamtenergie  $E$  des betrachteten physikalischen Systems und rechts die einzelnen Energieformen, wie mechanische Energieformen (Bewegungsenergie,  $\underline{v} \cdot d\underline{p}$ , Rotationsenergie,  $\underline{\omega} \cdot d\underline{L}$ , etc.), Wärmeenergie,  $TdS$  oder chemische Energie,  $\sum_j \mu_j dn_j$ , etc., die das betrachtete System mit einem anderen System austauscht, und die das betrachtete System auszeichnen. Betrachtet wird die Veränderung der Energie von einem Zustand 1, der durch einen konstanten Wert von  $E=E_1 (=const.)$  und konstante Werte der korrespondierenden Variablen  $\xi_{i,1}, X_{i,1}$  gegeben ist, und einem Zustand 2, dessen Energie  $E_2$  und die Werte der korrespondierenden Variablen  $\xi_{i,2}, X_{i,2}$  wiederum einen konstanten Wert besitzen. Tauscht das System keine Energie aus (Geschlossenheit), d.h. sind alle Werte  $dX_i$  auf der rechten Seite gleich null (d.h.  $X_i=const.$ ), dann ist das System für eine physikalische Betrachtung relativ bedeutungslos. Es befindet sich in einem Zustand, bei dem der Wert der Energie und alle das System aus physikalischer Sicht charakterisierenden Größen (definiert durch die Variablenpaare  $\xi_i, X_i$ ) einen konstanten Wert besitzen. Das System verharrt in diesem Zustand solange bis ein Energieaustausch mit einem anderen System stattfindet. Würde an dem System keine Energie ausgetauscht, dann verharrt es bis in alle Ewigkeiten in einem Zustand, dessen charakteristische Variablen ohne Energieaustausch auch nicht bestimmbar wären – mit anderen Worten, es wäre dann sinnlos von einem physikalischen System zu sprechen.

Wie man sieht, sind geometrische Systemabgrenzungen hierbei völlig überflüssig. Die Systemdefinition ist in der Physik durch eine abstrakte mathematische Beschreibung gegeben, in der Begriffe wie 'offen' oder 'geschlossen' keinen Platz haben. Das "System existiert", d.h. die rechte Seite von Gl.(7) ist ungleich null oder das "System existiert aus formaler Sicht nicht", d.h. die rechte Seite der Gl.7 ist null und somit  $dE=0$ ; daran ändern auch noch so spitzfindige philosophische Argumente nichts. Wie obsolet geometrisch-räumliche Abgrenzungen in der Physik geworden sind, geht schon aus der Atomphysik hervor. In diesem Zusammenhang sei nur an die Heisenberg'sche Unschärferelation erinnert.

- **Zusammenfassend lässt sich festhalten:**

Zur vollständigen Beschreibung des Zustandes eines physikalischen Systems gehört neben der Angabe der Werte verschiedener für das System (von einem Beobachter festgelegten) charakteristischer physikalischer Variablen immer die Angabe des Wertes seiner Energie.

Ein physikalisches System ist definiert durch die Angabe des Wertes seiner Energie und der komplementären Variablenpaare  $\xi_k, X_k$  die für eine adäquate Systembeschreibung vom Experimentator als notwendig erachtet und von ihm festgelegt werden. (8)

Der Begriff des 'Gleichgewichts' bezieht sich immer auf einen physikalischen Zustand, stellt also begrifflich nichts Neues dar. Befindet sich ein physikalisches System im Gleichgewicht, dann tut sich aus physikalischer Sicht nichts an dem System. Lebende Systeme sind aus physikalisch-chemischer Sicht immer "fernab" vom physikalisch-chemischen Gleichgewicht, Wären sie im physikalischen Gleichgewicht, dann wären sie im sprichwörtlichen Sinne "mausetot" und damit auch nicht zur Kognition befähigt.

**Lebende Systeme** werden permanent von Energie durchströmt, d.h. in sie strömt sowohl Energie hinein wie auch heraus. Was dagegen niemals aus ihnen heraus oder in sie

hineinströmt, ist Information und zwar auch dann nicht, wenn Elektroden im Gehirn angelegt werden. Es strömt immer nur Energie, das gilt auch für den Wahrnehmungsapparat oder die vom Experimentator im Gehirn angelegten Elektroden – hier könnte man allenfalls in der Sprache der Nachrichtentechnik von Signalen sprechen, die erst in dem empfangenden System selbst und/oder im Kopf des Experimentators: (Shannon'sche Informationstheorie) zur Information für das System und/oder den Experimentator werden. Dabei ist die Bedeutung der Information für das System nicht notwendigerweise identisch mit der Bedeutung der Information die "im Kopf des Experimentators errechnet" wird – das ist das Problem.

## 2. Die Reduktion biologischer Systeme zu "offenen" Systemen oder der Reduktionismus in der Biologie

Da durch jede Messung die Veränderung zwischen einem Anfangs- und einem Endzustand bestimmt wird, reduziert jede experimentelle Wissenschaft, bei der die Messung im Vordergrund steht, ein System zu einem offenen (Teil-)System, bei dem es einen 'Anfang' und eine 'Ende', einen 'Input' und einen 'Output' gibt. Das sind Begriffe, die nur im Zusammenhang mit offenen Systemen einen Sinn ergeben, und damit wird auch klar, warum in den klassischen Naturwissenschaften der Begriff des 'Systems' im allgemeinen nur eine untergeordnete Rolle spielt, von 'Geschlossenheit' zu sprechen ist in diesem Kontext sinnlos. Das gilt auch für die Biologie, die, wenn überhaupt von Systemen gesprochen wird, in aller Regel auf die aus formaler Sicht obsoleten räumlich-geometrischen Vorstellungen der Physik und Chemie zurückgreift.

Nun wird aber von Seiten der modernen Kybernetik 'Geschlossenheit' eines Systems für die Existenz einer 'Umgebung' und diese wiederum für die Beschreibung 'kognitiver' Prozesse gefordert, und es sind gerade die kognitiven Fähigkeiten, die lebende Systeme von toter Materie signifikant unterscheiden. Damit stellt sich diese Thematik nicht nur den Ingenieur- und Computerwissenschaften bei ihren Bemühungen um eine Modellierung und Simulation kognitiver Prozesse, sondern sie stellt sich auch und gerade den Biowissenschaften.

So reduziert sich aus wissenschaftlich konzeptioneller Sicht das System 'Affe', bei dem beispielsweise die Hirnaktivität als Funktion (vom Experimentator) vorgegebener äußerer optischer Reize durch Elektroden gemessen wird, auf das System eines lebenden, nicht-trivialen Signal- oder Datenfilters, bei dem ein lebendes neuronales Netzwerk eingesetzt wird. Durch die experimentelle Anordnung ist das System 'Affe' bzw. dessen Gehirn (für den Experimentator) zu einem offenen System reduziert worden. Über den (visuellen) Wahrnehmungs- oder Kognitionsprozess, der sich im System 'Affe' während der experimentellen Situation abspielt, erfährt der Experimentator durch solche oder ähnlich durchgeführten Messungen nichts. An dieser Situation würde sich natürlich auch dann nichts ändern, wenn experimentell die Möglichkeit bestünde, die Aktivität jedes einzelnen Neurons bis ins letzte Detail vermessen zu können. Erfolgreicher dagegen sind derartige Experimente, wenn es sich beispielsweise um Untersuchungen des Stoffwechsels handelt, dann befindet man sich jedoch eindeutig im Kontext von Physik und Chemie, und das ist ganz offensichtlich eine andere Beschreibungsdomäne als die der kognitiven Prozesse.

- Wo liegen die **epistemologischen Schwierigkeiten** bei dem Experiment der Messung von Hirnaktivitäten im Kontext einer wissenschaftlichen Beschreibung kognitiver Prozesse ?

- a) Da ist zunächst die heute noch immer gebräuchliche, jedoch wissenschaftlich völlig einseitige Verwendung des Begriffs von 'Information', wie er 1948 von *Shannon* eingeführt wurde, zu nennen. In der von *Shannon* entwickelten Informationstheorie wird der Informationsgehalt von Signalen in Analogie zu physikalischen Objekten als eine messbare Größe angesehen. Die Nützlichkeit dieser ausschließlich objektbezogenen Konzeption von Information für die Daten- und Signalübertragung in der Nachrichtentechnik ist unbestritten. Aus der Sicht einer modernen Kybernetik, die sich insbesondere auch mit der Beschreibung lebender Systeme beschäftigt, reicht diese Begriffsbildung jedoch nicht mehr aus, denn ein Signal wird erst in einem von dem empfangenden System (selbst) festgelegten Kontext zur Information für das betreffende System, d.h. Information existiert nicht *sui generis*.

Solange sowohl Biologen als auch (Neuro-)Informatiker diesen relativ einfach nachzuvollziehenden Sachverhalt ignorieren, werden ihren Bemühungen auf dem Weg zu einer Theorie kognitiver Prozesse keine durchschlagenden Erfolge beschieden sein.

- b) Die "Festlegung" physikalischer Systeme als "offene Systeme" ist natürlich keine bloße Marotte der Physiker, sondern entspringt dem Bemühen, physikalisch-chemische Systeme und Prozesse auch mathematisch beschreiben zu können. Dabei stellt die Mathematik als formale Sprache ein extrem effizientes Hilfsmittel für die wissenschaftliche Kommunikation dar. Die formale Beschreibung eines geschlossenen Systems im Sinne der 'Closure Thesis' (vgl. Aussage 3) ist auf der Basis der klassischen Mathematik nicht möglich.<sup>6, 7</sup>

An dieser Stelle taucht sofort die Frage auf, wie technische Systeme mit kognitiven Fähigkeiten konstruiert werden sollen, wenn diese Prozesse mathematisch nicht darstellbar sind? Hier offenbart sich ein Problem, welches sowohl die Fundamente der KI-Forschung, der Neuroinformatik aber auch der Robotik tangiert – es zu ignorieren, löst das Problem nicht.

Bis heute ist diese Erkenntnis der Nicht-Formalisierbarkeit kognitiver Prozesse mit den Mitteln der klassischen Mathematik sowie die sich daraus ergebenden wissenschaftlich-technischen Konsequenzen von Seiten dieser Disziplinen noch nicht einmal im Ansatz erkannt worden. Daraus resultiert auch der ungebrochene Glaube, man werde mit Hilfe der Modelle neuronaler Netzwerke und/oder ihrer Kombination mit den Methoden der Fuzzy-Logik, wenn man denn nur die Rechenleistungen (massive Parallelität) genügend steigern kann, eines Tages zu kognitiven, und damit zu autonomen Systemen in der Technik gelangen; aus wissenschaftlicher Sicht ist diese Vorstellung naiv. Die Modelle der Neuroinformatik sind mit ihren Input- und Output-Schichten offene Systeme *par excellence*, und damit sind sie NON-KOGNITIV.

- c) Selbst wenn es gelingen sollte, die Aktivitäten aller Neuronen des Gehirns zu messen, so wäre dieser Erfolg mikro-elektro-mechanischer Experimentierkunst noch kein

<sup>6</sup> von Foerster, H.: Entdecken oder Erfinden – Wie lässt sie Verstehen verstehen? in: Einführung in den Konstruktivismus (H. Gumin & A. Mohler, eds.), Oldenburg Verlag, München, 1985.

In dieser Publikation wird das Problem der 'Geschlossenheit' bei kognitiven Prozessen an Hand einer unendlichen Rekursion verdeutlicht – eine unendliche Rekursion ist mit keiner Turing Maschine bearbeitbar und damit auch nicht mit Hilfe eines Computers berechenbar.

<sup>7</sup> von Goldammer, E. & Kaehr, R.: Poly-contextural modeling of heterarchies in brain functions, in: Models of Brain Functions (Cotterill, R.M.J., ed.), Cambridge University Press, 1989, p.483-497.

Garant für einen Erkenntnisfortschritt auf dem Gebiet der Kognition. Dies soll im folgenden am Modell eines finiten Automaten demonstriert werden.<sup>8</sup>

In der Abb.1 ist das Prinzip eines extrem einfachen Automaten dargestellt. Es soll angenommen werden, dass dieser Automat durch vier Werte (A,B,C,D) der Eingangsvariablen  $x$  und durch zwei Werte (0,1) der Ausgangsvariablen  $y$  charakterisiert sei. Zwischen den Werten der Eingangs- und Ausgangsvariablen existiere ein funktionaler Zusammenhang wie z.B. der in Tab.1 angegebene. D.h., liegt am Eingang die Sequenz (ABCD) an, dann wird die Sequenz (0110) durch den Automaten errechnet, die am Ausgang anliegt. Anders ausgedrückt, wann immer der Maschine als Eingangssignal (Ursache) das Symbol  $x=A$  angeboten wird, ist das Resultat ihrer Operation das Ausgangssymbol (Wirkung)  $y=0$ . Eine solche Maschine wird von *Von Foerster* als "Triviale-Maschine" bezeichnet, ihre Funktion in der Datenverarbeitung wäre z.B. die eines digitalen Filters, wie etwa die (adaptierten) neuronalen Netzwerke, die ebenfalls zur Klasse der trivialen Maschinen gehören. Die Adaptionsphase bei den neuronalen Netzwerken kann sozusagen als Trivialisierungsprozess dieser Netzwerkmodelle angesehen werden.

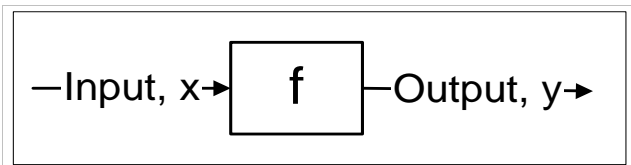


Abb.1 : Prinzip eines "Trivialen Automaten"

f	
x	y
A	0
B	1
C	1
D	0

Tabelle\_1 : Funktionaler Zusammenhang  $y = f(x)$  für einen "Trivialen Automaten"

Triviale Automaten sind: 1) synthetisch determiniert; 2) analytisch determinierbar; 3) vergangenheitsunabhängig; und 4) vorhersagbar. Automaten dieser Art sind nicht nur langweilig, mit ihnen könnte man noch nicht einmal einen herkömmlichen Computer bauen.

In der Abb.2 ist ein Automat angegeben, der sich von seinem trivialen Pendant dadurch unterscheidet, dass die Operationen dieser Maschine von den jeweiligen "inneren Zuständen"  $z$  der Maschine abhängen. Der Maschinentyp wird im folgenden mit dem Akronym NTA (Nicht-Trivialer Automat) abgekürzt. Betrachtet sei wiederum der einfachste Fall eines NTA mit nur zwei internen Zuständen I und II, wie dies in der Tab.2 angezeigt ist.

Wird an den Eingang (Ursache) das Signal  $x=A$  angelegt, so ist das Resultat der Operation (Wirkung)  $y=0$ . Eine Sequenz (A,A,A,...) ergibt die Sequenz (0,0,0,...) am Ausgang. Entsprechend ergibt die Sequenz (B,B,B,...) als Resultat die Sequenz (1,1,1,...) am Ausgang. Will man nun wissen, was C bewirkt, d.h. wird an den Eingang das Signal (Ursache)  $x=C$  angelegt, dann ist das Resultat der Operation (Wirkung)  $y=1$ . Die Wiederholung dieses Vorgangs (erneutes Anlegen von C an den Eingang) liefert jedoch  $y=0$ , da der Automat beim ersten Anlegen von C vom Zustand I in den Zustand II übergegangen ist. Wird jetzt an den Eingang wiederum  $x=B$  angelegt, dann ist das Resultat in diesem Fall  $y=0$  und die Sequenz (B,B,B,...) würde jetzt zu (0,0,0,...) führen, da sich der Automat jetzt im Zustand II befindet. Wird  $x=D$  angelegt, dann ergibt sich  $y=1$ . Eine Sequenz (D,D,D,...) ergibt, wie man der Tab.2 entnehmen kann, die Sequenz (1,0,1,0,...) bzw. wenn vom Zustand I aus gestartet wird, (0,1,0,1,...).

<sup>8</sup> siehe dazu auch Ref. /6/.



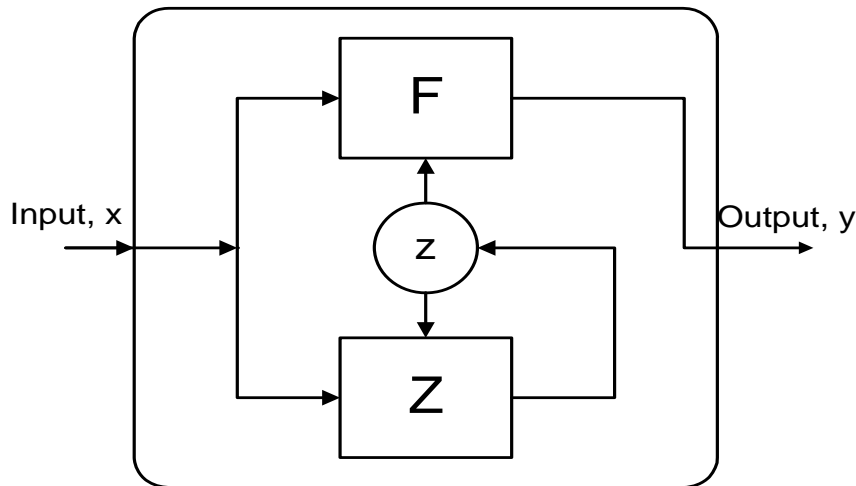


Abb. 2 : Prinzip eines "Nicht-Trivialen-Automaten"

f					
Zustand I			Zustand II		
x	y	z'	x	y	z'
A	0	I	A	1	I
B	1	I	B	0	II
C	1	II	C	0	II
D	0	II	D	1	I

Tabelle 2: Funktionaler Zusammenhang  $y = f(x; z)$  für einen "Nicht-Trivialen Automaten"

Angenommen der NTA stellt ein zu untersuchendes unbekanntes Objekt dar, dann wird ein Experimentator, der den funktionalen Zusammenhang  $y=F(x,z)$  und  $z'=Z(x,z)$  aus der Tab.2 durch wiederholtes Experimentieren erst finden soll, diesen Zusammenhang zwischen  $x$ ,  $y$  und  $z$  nach einer Reihe von Versuchen möglicherweise finden. Die interessante Frage ist, wie groß ist die Anzahl der möglichen NTA's, die durch jeweils eine Tabelle (entsprechend der Tab.2) repräsentiert werden?

Nehmen wir an, dass sich die Eingangsgrößen durch zweistellige Binärzahlen darstellen lassen, also  $A=00$ ,  $B=01$ ,  $C=10$ ,  $D=11$ , dann gibt es für diesen relativ einfach Fall bereits insgesamt  $6 \cdot 10^7$  verschiedene Kombinationen, wobei jede Kombination einem NTA bzw. einer Tabelle entspricht (siehe Ref.6). Die Anzahl der möglichen NTA's wächst astronomisch an, wenn sich die Zahl der Eingangsvariablen und/oder der Zustände erhöht.<sup>9</sup>

Maschinen dieser Art werden in Ref.6 als 1) synthetisch determiniert, 2) analytisch unbestimmbar, 3) vergangenheitsabhängig und 4) unvoraussagbar bezeichnet.

Es ist in diesem Kontext nicht weiter relevant nach dem wissenschaftlichen Nutzen von Messungen der Hirnaktivität bei Tieren zu fragen, das steht hier nicht zur Debatte. Fest steht jedoch, dass auf diese Weise keine Erkenntnisse über den Prozess der Kognition gewonnen werden können, der Vergleich mit einem NTA liegt auf der Hand. Wenn aber die Fähigkeit

<sup>9</sup> Unbestimmbarkeitsprinzip nach A. Gill: Gill hat gezeigt, dass es funktionale Organisationen solcher finiter Automaten gibt, die prinzipiell nicht durch eine endliche Versuchsfolge erschlossen werden können: Gill, A., Introduction to the Theory of Finite-State-Machines, Mc-Graw-Hill, N.Y., 1962.

zur Kognition ein wesentliches Merkmal lebender Systeme ist, dann tragen Experimente dieser Art, wenn überhaupt, nur wenig zu einer "Theorie lebender Systeme" bei und damit nützen sie einem Ingenieur, der kognitive Fähigkeiten in einem technischen System abbilden möchte, nichts.

An dieser Stelle ist es angebracht, sich in einem Gedankenexperiment folgenden hypothetischen Fall vorzustellen: Angenommen, einem Ingenieur sei die Konstruktion eines technischen Systems mit kognitiven Fähigkeiten gelungen – wie will die "Nachwelt" den Algorithmus eines solchen Systems erforschen, wenn dieser nicht publiziert wurde?

Durch Messungen der elektrischen Signale, d.h. deren Amplituden und/oder Frequenzen, gelingt dies aus den oben dargelegten Gründen ebenso wenig, wie es einem Biologen gelingt, aus elektrophysiologischen Messungen der Hirnaktivität etwas über den Prozess der Kognition zu erfahren. Schlimmer noch, selbst der Algorithmus eines einfachen Programms, wie es heute in jedem gewöhnlichen Computer abläuft, lässt sich auf diesem Weg nicht finden (vgl. dazu Ref.9).

### Fassen wir kurz zusammen:

- Die Systemdefinition in der Physik erfolgt über mathematische Differentialgleichungen. Alle räumlich-geometrischen Systemdefinitionen sind spätestens seit Einführung der Quantenmechanik in die Physik völlig obsolet geworden.
- Das komplementäre Begriffspaar 'offen' und 'geschlossen' hat in einer physikalischen Theorie, in der messbare Größen dominieren, keinen Platz.
- Kognitive Prozesse gehören in eine andere Domäne der Beschreibung und nicht in die Beschreibungsdomäne physikalisch-chemischer Prozesse.
- Benutzt man für die Beschreibung kognitiver Prozesse den Sprachrahmen, der durch die klassische (zweiwertige) Logik aufgespannt wird, dann führt die Beschreibung kognitiver Prozesse zu Systemen bzw. Modellen, die als 'geschlossen' im Sinne eines logischen *circulus vitiosus* angesehen werden müssen. Der Versuch ihrer Beschreibung auf der Basis offener Systeme oder Modelle ist identisch mit einer Systemreduktion (Reduktionismus); nur in diesem Falle sind die Kalküle der klassischen (mono-kontexturalen) Logikkonzepte widerspruchsfrei anwendbar<sup>10,11,12</sup> – der zu beschreibende Prozess jedoch ist für das formale Modell verloren gegangen.
- Das Problem das es zu lösen gilt, ist seit mehr als 2000 Jahren unter verschiedenen Etiketten wie die 'Dichotomie von Geist und Materie' oder die 'Subjekt-Objekt-Spaltung' bekannt. Heute tritt es als das Problem einer Verknüpfung verschiedener Beschreibungsdomänen erneut auf, nämlich einmal der physikalisch-chemischen Prozesse auf der einen Seite und der kognitiven Prozesse auf der anderen Seite.

---

<sup>10</sup> Kaehr, R. & von Goldammer, E.: Again Computers and the Brain, Journal of Molecular Electronics, Vol. 4, 1988, p.S31-S37.

<sup>11</sup> von Goldammer, E. & Kaehr, R.: Problems of Autonomy and Discontextuality in the Theory of Living Systems, in: Analyse dynamischer Systeme in Medizin, Biologie und Ökologie, Reihe: Informatik-Fachberichte (Möller, D.P.F. & Richter, O., Hrsg.), Springer Verlag, 1990, p.3-12.

<sup>12</sup> von Goldammer, E. & Kaehr, R.: Lernen in Maschinen und lebenden Systemen, Design und Elektronik, März 1989, p.146-151.

- Die Entwicklung einer vereinheitlichten Theorie beider Beschreibungsdomänen ist primär ein wissenschaftslogisches und kein experimentelles Problem, worauf in der Vergangenheit schon mehrfach hingewiesen wurde – siehe dazu Ref.10-12.

### III. Autonome Systeme: Beschreibung und Konstruktion

Sowohl die Modellierung autonomer Systeme, deren Simulation sowie die Formulierung einer Theorie lebender Systeme stellen eine äußerst komplexe interdisziplinär orientierte Aufgabe dar, die, ebenso wie die Konstruktion autonomer Fahrzeuge, d.h. zur Kognition befähigter technischer Artefakte, eine (formale) Sprache als Kommunikationshilfe und Konstruktionsgrundlage voraussetzt. Eine adäquate formale Sprache muss, wie in Abschnitt II dargelegt wurde, so strukturiert sein, dass verschiedene Beschreibungsdomänen durch dafür geeignete Operatoren miteinander vernetzt werden können, und nicht wie bisher beziehungslos irgendwie nebeneinander stehen. Das ist sozusagen die minimale Voraussetzung für eine vereinheitlichte Theorie physikalischer (und chemischer) mit kognitiven (und volitiven) Prozessen, wie sie nun einmal in lebenden Systemen ganz offensichtlich auftreten.

Da alle wissenschaftlichen Aussagen logisch fundiert sein müssen, ist als formale Sprache zunächst ein Logik-Kalkül gefordert, bei dem verschiedene logische Domänen<sup>13</sup> durch geeignete Operatoren bereits so miteinander vernetzt sind, dass eine Beschreibung, d.h. Modellierung und Simulation, simultan-parallel ablaufender physikalisch-chemischer und kognitiv-volitiver Prozesse möglich wird.

Ein derart parallel vernetzter Kalkül ist durch die **polykontexturale Logik** (PKL) und die ihr zugrunde liegende Stellenwerttheorie, der Keno- und Morphogrammatik bereits gegeben. Das Problem einer formal wissenschaftlichen Beschreibung lebender Systeme in einem "ganzheitlichen" Sinne besteht in der Aufgabe, diesen Kalkül in adäquater Weise mit Semantiken zu füllen. Dies stellt die bereits erwähnte komplexe interdisziplinäre Aufgabe dar, für die es keine historischen Vorbilder gibt. Wissenschaftsgeschichtlich gesehen ist dies eine völlig neue Situation, denn in der Vergangenheit wurde parallel zur Entwicklung der begrifflichen Fassung einer Theorie, die aus der experimentellen Beobachtung abgeleitet wurde, ein Formalismus entwickelt. Die mathematisch formale Weiterentwicklung einer solchen Theorie führte dann zur Reduktion der Bedeutungsvarianz der verwendeten Begriffe.<sup>14</sup>

Eine weitere grundsätzliche Schwierigkeit auf dem Weg zu einer '**Theorie lebender Systeme**' resultiert aus der Tatsache, dass wir nicht über die Fähigkeit verfügen, Begriffe simultan-parallel zu denken, schlimmer noch, eine derartige Parallelität von Prozessen kann weder gemessen noch sonst in irgendeiner Weise unmittelbar wahrgenommen werden. Das bedeutet jedoch nicht, dass solche Prozesse möglicherweise gar nicht existieren – für das Verständnis lebender Systeme sind sie von fundamentaler Bedeutung. Erst die Begrifflichkeit einer derartigen Prozessualität, bei der Probleme wie z.B. **Mehrzeitigkeit, Po-**

---

<sup>13</sup> In einer logischen Domäne sind jeweils alle logischen Operationen gültig.

<sup>14</sup> So versteht man heute unter einem physikalischen "Teilchen" den Energie- und Impulstransport durch den freien Raum. Damit hat sich die Bedeutung des Begriffes "Teilchen" von etwas Greifbarem ins Abstrakte hin gewandelt. Dies ist eine unmittelbare Folge der Atomphysik, in der z.B. Licht sowohl als Welle wie auch als Teilchen (Photon) beschrieben werden kann. Bei einem Billardspiel sind nicht die greifbaren Kugeln von physikalischem Interesse, sondern wiederum nur der Energie- und Impulstransport sowie deren Austausch beim Stoss.

**lyrhythmie**, etc. auftauchen und theoretisch zu bewältigen sind, führen zu einer 'Theorie des Lebens'.

Alle heute bekannten parallelen Rechnerarchitekturen und die auf diesen Plattformen ablaufenden Algorithmen lassen sich grundsätzlich auch sequentiell abarbeiten, ohne dass der durch sie abgebildete Prozess eine qualitative Veränderung erfährt – was sich dabei verändert ist die Bearbeitungsgeschwindigkeit des Algorithmus. Diese Art von Parallelität ist nicht gemeint, wenn hier von simultaner Parallelität die Rede ist.

Da menschliches Denken aus inhaltlicher Sicht, d.h. in Sprache und/oder Bildern sequentiell abläuft, der Gesamtprozess des Denkens selbst jedoch ein simultan-parallel ablaufender Prozess ist, der sich prinzipiell nicht sequentiell abbilden und daher auch nicht messen lässt, ist für die Entwicklung einer 'Theorie des Lebens' der Computer ein unverzichtbares Hilfsmittel (Modellierung) und Werkzeug (zur Simulation). Der Weg zu einer 'Theorie des Lebens' führt sowohl methodisch als auch methodologisch über die Computerwissenschaften im Sinne einer **Symbiose von Computer- und Biowissenschaften**.

Obwohl die Bezeichnung 'Neuroinformatik' den Eindruck erweckt, als wäre durch die Renaissance dieses Gebietes zu Beginn der 80er Jahre eine derartige Symbiose bereits erfolgt, trägt der Schein, wie dies auch aus den Argumenten des Abschnitts II hervorgeht. Berücksichtigt man, dass einige Modelle der Neuroinformatik in den Lehrbüchern der Physik bzw. der Elektrotechnik unter anderen Bezeichnungen abgehandelt werden, wie beispielsweise als nicht-lineare Signal- und Daten-Filter und als solche seit Jahren bekannt sind und technisch erfolgreich appliziert werden, dann wird wiederum deutlich, dass es auf diesem Weg zu einer wirklichen Symbiose von Bio- und Computerwissenschaften kaum kommen kann. So stellt beispielsweise das Paradigma der Neuroinformatik (bzw. des Neo-Konnektionismus) für die Psychologie einen Rückfall in den Behaviorismus dar, der heute unter Psychologen als überholt angesehen wird. Das Problem wird noch deutlicher, wenn man sich vergegenwärtigt, dass die adaptierten Netzwerkmodelle zur Kategorie der 'trivialen Automaten' gezählt werden müssen, was aus Sicht der Lernpsychologie allenfalls der "Konzeption des Nürnberger Trichters" entsprechen würde, wenn es denn eine solche überhaupt gibt.

Interessanter und vielleicht auch wegweisender sind in diesem Zusammenhang Entwicklungen anzusehen, die unter der Bezeichnung 'Artificial Life'<sup>15</sup> oder 'Multiple Agents'<sup>16</sup> bekannt wurden sowie alle Bemühungen seitens der modernen Robotik, autonome Systeme zu konzipieren.<sup>17</sup> Hier wird zumindest der Versuch unternommen, Eigenschaften zu simulieren und nachzubilden, die lebende Systeme auszeichnen. Auch wenn man diese Experimente noch als sehr rudimentär und aus wissenschaftlicher Sicht teilweise vielleicht sogar als naiv ansehen muss, so sorgen zumindest alle jene Ansätze, die zu technisch-wirtschaftlichen Umsetzungen führen, wie die der Robotik, für den notwendigen Druck, um eine Umsetzung des eingangs erwähnten Paradigmenwechsels in den (Bio-)Wissenschaften zu beschleunigen, wo diese Entwicklungen bisher kaum zur Kenntnis genommen wurden. Im folgenden soll die Thematik autonomer Systeme im Zusammenhang mit der Robotik etwas näher beleuchtet werden.

---

<sup>15</sup> Langton, C.G., (ed.), *Artificial Life*, Addison-Wesley Publ., 1989.

<sup>16</sup> Müller, J. (Hrsg.), *Verteilte Künstliche Intelligenz*, BI-Wissenschaftsverlag, Mannheim, 1993.

<sup>17</sup> Antsaklis, P.J. & Passino, I.K.M. (eds.), *An Introduction to intelligent and Autonomous Control*, Kluwer Academic Publ. 1993.

## 1. Konzeption eines adaptiven Reglers

### – supervised & unsupervised learning –

In der Regelungstechnik besteht ein Problem darin, geeignete Regler so zu entwerfen, dass sie nicht nur sehr allgemein anwendbar, sondern vor allem stabil und robust sind. Dabei kommen heute auch neuronale Netzwerke (NN), Fuzzy-Regler (FR) und die Kombination aus beiden, die Neuro-Fuzzy-Systeme (NFS) zur Anwendung. Der Vorteil dieser "unkonventionellen" Methoden liegt vor allem darin, dass ohne die Kenntnis des genauen mathematischen Modells einer Regelstrecke bzw. der mathematischen Beschreibung von Mustern oder Objekten diese mit Hilfe von FR- und/oder NN-Techniken erfasst und als 'Look-Up-Table' für Echtzeit-Anwendungen vorgehalten werden können. An einem Beispiel soll das Problem verdeutlicht werden.

#### BEISPIEL: "UNTERWIESENES\_und\_NICHT\_UNTERWIESENES LERNEN"

Ein Roboter soll von Punkt A einen Gegenstand nach Punkt B transportieren (z.B. in einem Krankenhaus das Essen oder in einem Hotel das Gepäck zwischen der Rezeption und den Zimmern verteilen). Dies Beispiel gehört nicht dem Bereich der Science Fiction an, solche Systeme gibt es bereits.

In der Abb.3 ist die Situation für die folgende Diskussion nochmals etwas vereinfacht skizziert. Ein Fahrzeug soll in einem begrenzten Raumbereich selbständig auf einen Parkplatz P fahren. Dazu sei das Fahrzeug mit Sensoren ausgestattet, die auf einer Achse befestigt sind und von einem Schrittmotor um definierte Winkel gedreht werden können. Damit lässt sich die Position des Fahrzeugs bestimmen, die durch die Variablen  $x$ ,  $y$  und den Winkel  $\psi$  repräsentiert werden. Das Fahrzeug sei mit einer eigenen Recheneinheit zur Auswertung der Sensordaten sowie zum Ansteuern der Aktoren (Schrittmotor-Lenkung/Winkel-Vorderrad; Motor/Geschwindigkeit vor und zurück) ausgestattet. Das Fahrzeug ist ferner über eine Funkfernsteuerung mit einem externen Rechner bzw. einem Operateur verbunden.

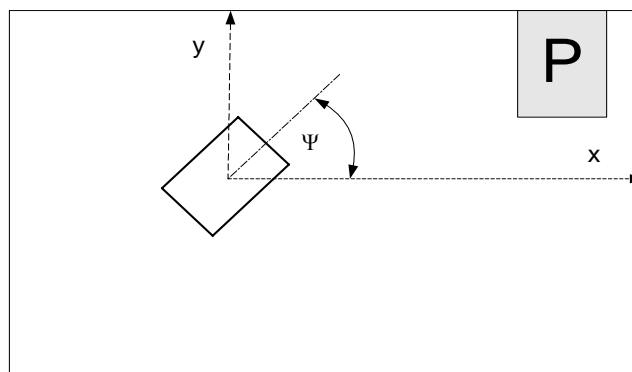


Abb. 3 : Regelungskontext "Einparken in eine vorgegebene Position"

In der 1. Phase (UNTERWIESENES LERNEN) wird das Fahrzeug von unterschiedlichen Startpositionen aus über die Funkfernsteuerung vom Operateur auf die Parkfläche gesteuert. Die über die Sensoren jeweils ermittelten Positionen werden zusammen mit den korrespondierenden Aktordaten registriert. Sie bilden die Datensätze für den Regler, die in einer 'Look-Up-Table' verwaltet werden: Eingangsdaten:  $x, y, \psi$ ; Ausgangsdaten: Einstellwinkel-Vorderrad, Geschwindigkeit (Betrag vorwärts bzw. rückwärts).

In der 2. Phase (NICHT-UNTERWIESENES LERNEN) findet das Fahrzeug aus allen möglichen Positionen selbständig den Platz P, d.h. auch aus Positionen, die vorher nicht

eintrainiert wurden. Die 'Look-Up-Table' wird sozusagen ergänzt. Darüber hinaus soll das Fahrzeug die Anfahrwege, die Trajektorien, zum Ort P aus eigener Leistung nach vorgegebenen Kriterien optimieren.

Letzteres wird in einer 3. Phase der Adaption von Bedeutung, wenn ein vorher nicht vorhandenes Hindernis auf der Fahrt zur Parkposition P eingebracht wird, und dieses Hindernis wiederholt angetroffen und umfahren werden muss. In diesem Fall ist es zweckmäßig, das Hindernis als Objekt in die 'Look-Up-Table' mit aufzunehmen, um den Fahrtweg zu optimieren. Hier muss das System in der Lage sein, selbständig eine Entscheidung zu treffen, ob die 'Look-Up-Table' ergänzt werden soll oder nicht.

In diesem Beispiel wird im 1. Schritt das System in einer Anlernphase, die als 'Lernen\_0' bezeichnet werden soll, sozusagen vorprogrammiert. Konzeptionell gesehen entspricht diese Form des Lernens dem so genannten 'Supervised Learning', wie es im Sprachgebrauch des Konnektionismus benannt wird.

Aus dem Beispiel wird deutlich, dass in Schritt 2 und 3, also nach der Adaptionphase, mit einer 'Look-Up-Table' gearbeitet wird, die im 1. Schritt erstellt wurde. Die Veränderungen der Daten dieser 'Look-Up-Table' durch Ergänzung und/oder Optimierung sowie die Implementierung der Koordinaten möglicher Hindernisse in die 'Look-Up-Table' durch das System selbst wird im folgenden als 'Lernen\_I' bezeichnet. Konzeptionell entspricht dies dem 'Unsupervised Learning' der Neuroinformatik.

Ohne an dieser Stelle darauf einzugehen, wie die Phasen 1 bis 3 jeweils konkret realisiert werden können – dies ist keine triviale aber dennoch lösbare Aufgabe – ist im vorliegenden Zusammenhang der in allen drei Phasen zeitlich invariante, d.h. sich nicht verändernde Kontext des Prozesses entscheidend, nämlich "Einparken in eine vorgegebene Position". Prozesse dieser Art besitzen sozusagen fließbandähnlichen Charakter, d.h. sich ständig wiederholende gleiche Situationen. Dies gilt auch für das neu aufgetauchte Hindernis aus Phase 3, das als eine Störung oder "starkes Rauschen" angesehen werden kann, bei dem der globale Kontext des Prozesses "Einparken-in-eine-vorgegebene-Position" trotz Störung erhalten bleibt.

Das Beispiel steht exemplarisch für sehr viele Regelungsaufgaben, bei denen die klassischen Methoden der Regelungstechnik ihre Anwendung finden. Ob dabei neuronalen Netzwerken, Fuzzy-Reglern oder Kombinationen davon der Vorzug gegeben wird, ist ausschließlich eine Frage der Zweckmäßigkeit, d.h. man wird die flexiblere, robuste und vor allem die am einfachsten zu realisierende und damit die wirtschaftlich günstigste Lösung wählen. Mit anderen Worten, Prozesse, die hier mit dem Etikett des Lernen\_0 bzw. Lernen\_I gekennzeichnet wurden, lassen sich sowohl mit den herkömmlichen Regelalgorithmen wie auch mit den NN-, FR- oder NFS-Techniken erfolgreich realisieren.

## 2. Lernen zu lernen oder Lernen\_II

Betrachtet man das Beispiel aus dem vorstehenden Abschnitt III.1 als eine Vorstufe für den Antrieb eines Haushaltsroboters, dann ist eine Regelung auf der Basis von Lernen\_0 und Lernen\_I ganz offensichtlich noch nicht ausreichend. Um dies einzusehen, stellen wir uns vor, die Parkposition im obigen Beispiel sei durch ein Hindernis derart blockiert, dass das Fahrzeug das Hindernis nicht umfahren und somit die Parkposition nicht erreicht werden kann. Wenn der oben beschriebene Regler technisch vernünftig konzipiert ist, dann wird das Fahrzeug stoppen und eine Fehlermeldung anzeigen. Niemand wird erwarten, dass unter den gegebenen Voraussetzungen das Fahrzeug – aus eigener Leistung – den Versuch unter-

nimmt, das Hindernis beiseite zu schieben oder eine neue Parkposition zu definieren, es sei denn, eine solche Option wurde vorher einprogrammiert.

Selbstverständlich ist es möglich, eine derartige Option vorher einzuprogrammieren. Dies bedeutet lediglich, dass der Kontext, in dem sich der betrachtete Prozess abspielen soll, durch diese Veränderung des Algorithmus entsprechend erweitert wurde, d.h. für das betrachtete Beispiel 'Einparken in eine vorgegebene Position und wenn nötig ein Hindernis beiseite räumen oder eine neue Position definieren'. Eine strukturell neue Lernsituation wird dadurch nicht erzeugt und deshalb ist diese Option für die weitere Diskussion uninteressant.

Stattdessen fordern wir von dem System, dass es auf ein derart unerwartetes Ereignis aus eigener Leistung eine Erweiterung des Kontextes vornimmt. Das bedeutet, dass das System in der Lage sein muss, den eigenen Algorithmus zu verändern und nicht nur den Datensatz wie bei Lernen\_I. Genau das würde man im alltäglichen Leben unter 'Lernen' in einer entsprechenden Situation verstehen. Bei diesem Lernen ändert sich das Verhältnis zwischen dem lernenden System und seiner Umgebung, und das stellt technisch gesprochen eine Veränderung des Algorithmus dar, der das betreffende System charakterisiert.

Ein System, welches dazu in der Lage ist, muss über **kognitive und volitive** Fähigkeiten verfügen. Es muss eine neue Situation wahrnehmen, sie reflektieren und entscheiden, d.h. einen Kontext festlegen, in dem die empfangenen Signale eine Bedeutung für das System erhalten. Mit anderen Worten, der Prozess 'Lernen-eines-Systems-mit-Umgebung', d.h. Lernen\_II umfasst wenigstens zwei sich wechselseitig bedingende, simultan-parallel ablaufende Prozesse (vgl. Ref.7):

- (i) ☞ ... a volitive (decision making) process structuring the environment by a determination of relevances and a corresponding context of significance within the semantical domain produced by (ii) ...
- (ii) ... a classification and abstraction of the data by cognitive processes producing (9) a representational structure of content and meaning within the context chosen in (i) ☞ ...

Beide Prozesse sind komplementär zueinander, d.h. sie bedingen sich gegenseitig und es macht keinen Sinn sie einzeln, d.h. voneinander unabhängig, betrachten zu wollen. Da unsere Vorstellungen heute sehr stark durch eine physikalische Begrifflichkeit und Anschauung geprägt ist, werden diese Prozesse häufig als getrennt ablaufend betrachtet. Damit wird aus logischer Sicht jedoch eine vollständig andere Prozessualität beschrieben. Der durch die Beziehung (9) gegebene Prozess lässt sich weder auf einen hierarchisch strukturierten Entscheidungsbaum abbilden, noch lässt er sich sequentiell darstellen oder beschreiben.

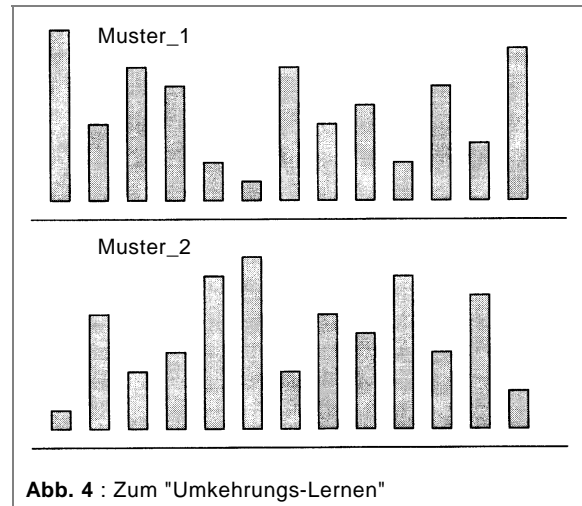
Die formale Darstellung derartiger Prozesse verlangt vom System die logische Unterscheidung zwischen einem Objekt (z.B. dem konkreten Hindernis) und dem Abbild des Objekts: Der Tisch, der vor einem steht, ist logisch gesehen von anderem Typus wie der Begriff des Tisches, also sein Abbild, oder um es in der Sprache der Mengenlehre auszudrücken, die Menge ist von logisch höherem Typus als ihre Elemente, gleiches gilt für das Verhältnis von Operator und Operand.

Ein System, welches lernt, wie es lernt, also der Prozess 'Lernen lernen', wird als Lernen\_II bezeichnet.<sup>18</sup> Die formale Darstellung eines derartigen Prozesses, bei dem das System die

<sup>18</sup> siehe Ref. /12/

gesamte Situation reflektiert und seinen eigenen Algorithmus verändern soll, erfordert demnach ein Vertauschen von Operator und Operand, d.h. was von einem Standpunkt ein Operator war, wird im Prozess der Reflektion zum Operanden und umgekehrt. Dieser Umtauschprozess muss simultan-parallel ablaufen. Das ist das Problem, welches der Aussage (9) zugrunde liegt und das es zu lösen gilt, wenn Prozesse wie 'Lernen II' technisch realisiert werden sollen.

Aus der Abb.4 lässt sich anhand zweier Muster der Prozess von Lernen\_II am so genannten "Umkehrungs-Lernen" nochmals verdeutlichen: Einem neuronalen Netz wird das Muster 1 angeboten, das Netz wird auf dieses Muster trainiert. Danach wird das Muster 2 angeboten und das Netz wird auf das Muster 2 adaptiert. Danach wird wiederum Muster 1 angeboten, dann wieder Muster 2 usw. Die entscheidende Frage lautet, was lernt das System (in diesem Fall das neuronale Netzwerk) aus eigener Leistung über die Umkehrung? Bei einem lernfähigen System im Sinne von Lernen\_II würde sich nach einiger Zeit die jeweilige Dauer der einzelnen Trainingsphasen verkürzen. Ein herkömmliches neuronales Netzwerk lernt aus eigener Leistung nichts über die Umkehrung des Adaptionsprozesses.



Obwohl jeder Kognitionsprozess ein Prozess '2. Ordnung' im Sinne von Lernen\_II ist, und obwohl lebende Systeme sich von toten Objekten gerade durch ihre Fähigkeiten zur Kognition signifikant unterscheiden, gibt es bisher weder eine technische Realisierung von Lernen\_II, noch gibt es bis heute technische Realisierungen kognitiv-volitiver Prozesse. Die Modelle der Neuroinformatik gehören zur Kategorie der Lernprozesse 1. bzw. 0. Ordnung und stellen konzeptionell gesehen, wie schon in Abschnitt II erwähnt, digitale Datenfilter, d.h. NON-KOGNITIVE Netzwerke dar.

- Die Frage stellt sich, "ob es überhaupt notwendig ist, derartige Prozesse formal zu beschreiben, um sie anschließend technisch zu realisieren?"

Dazu sei nochmals auf das Beispiel des parkenden Fahrzeuges in der Variante mit 'verbarriadiertem Parkposition' zurückgegriffen. Die gestellte Frage läuft technisch gesehen auf die Alternative hinaus, den Kontext für die Aktionen des Systems entweder durch vorherschauendes Programmieren adäquat zu erweitern, d.h. das vom Standpunkt des Systems unerwartete Ereignis wird vom Konstrukteur vorausschauend einprogrammiert, oder das System reagiert selbst auf das unerwartete Ereignis und modifiziert den eigenen Algorithmus (Lernen II), um auf diese Weise seinen Handlungskontext (selbst) zu erweitern. Dabei könnte das System zu unterschiedlichen Entscheidungen gelangen, nämlich eine neue Parkposition aufsuchen oder das Hindernis beseitigen. Beide Resultate stellen eine adäquate Erweiterung des globalen Handlungskontextes "Einparken-in-eine-vorgegebene-Position" dar.

Die Notwendigkeit der technischen Realisierung von Lernen\_II wird sofort deutlich, wenn man zu dem Szenarium 'Haushaltsroboter' übergeht. Die Vorstellung, jede mögliche Situation vorausschauend einprogrammieren zu können, ist absurd, insbesondere wenn ein sol-



cher Roboter in irgendeinem realen Haushalt und nicht in einer künstlichen Klötzchenwelt agieren soll. Eine flexible Signal- und Datenverarbeitung verlangt hier nach Prozessen wie Lernen\_II, das bedarf keiner weiteren Begründung. Prozesse von Lernen\_I alleine sind für die Konzeption eines Haushaltsroboters nur bedingt von Nutzen, denn es macht keinen Sinn, jede beliebige, neue Situation in eine 'Look-Up-Table' aufzunehmen. Genau hierin, nämlich in der Frage, ob eine neue Situation in den Datensatz eines globalen Kontextes übernommen werden soll oder nicht, muss vorher vom System selbst entschieden werden. Solche Entscheidungsprozesse werden aber erst durch die Realisierung von Lernen\_II ermöglicht.

### **Fassen wir kurz zusammen:**

- Lernen\_0 steht für Prozesse, bei denen sich weder der Algorithmus noch der Datensatz verändert. Der Datensatz ist entweder in der Form einer 'Look-Up-Table' oder in Silikon oder in beliebig anderer Hardware gespeichert.
- Lernen\_I steht für adaptive Prozesse, bei denen das System aus eigener Leistung den gespeicherten Datensatz, die 'Look-Up-Table', einer veränderten Situation anpasst, sie ergänzt oder zielorientiert (gemäß Regelungskontext) optimiert. Die Veränderungen sind dabei von 'lokaler' Natur, der 'globale Kontext', in dem das System agiert, bleibt ebenso wie der das System charakterisierende Algorithmus unverändert.
- Lernen\_II steht für adaptive Prozesse, bei denen nicht nur der Datensatz (die Operanden), sondern vor allem der Algorithmus (die Operatoren) vom System selbst verändert werden. Dies entspricht einer Erweiterung des Regelungskontextes, in dem das System agiert – das System lernt (aus eigener Leistung) etwas über seine Umwelt.
- Lernen\_0 und Lernen\_I lassen sich mit den Modellen der Neuroinformatik sowie mit Methoden der Fuzzy Logik, aber auch allen anderen mathematischen Hilfsmitteln nachbilden und technisch applizieren. In die Kategorie von Lernen\_I gehören auch die genetischen Algorithmen, die sich technisch zur Lösung von Optimierungs- und Klassifikationsaufgaben besonders bewährt haben.
- Kognitive Prozesse gehören in die Kategorie von Lernen\_II und wurden bis heute technisch noch nicht realisiert.

## **IV. RESÜMEE**

Der Stand wissenschaftlicher Erkenntnis stellt sich heute im Allgemeinen so dar, dass von Autonomie, Kognition, Lernen usw. im Kontext biologischer und technischer Systeme zwar gesprochen, die Problematik der Geschlossenheit jedoch nicht zur Kenntnis genommen wird. Dabei wird in aller Regel auch übersehen, dass autonome Systeme nicht nur über kognitive sondern auch über volitive Fähigkeiten verfügen müssen.<sup>19</sup> Damit nimmt die Komplexität des Problems der (formalen) Beschreibung autonomer Systeme noch erheblich zu.

Begriffe wie 'Offenheit' und 'Geschlossenheit' resultieren nicht aus experimentellen Messungen, sondern sind wie 'rechts' und 'links' standpunktabhängige Beschreibungskategorien, die erst im Rahmen einer formalen Darstellung ihre eigentliche Bedeutung erlangen. Alle räumlich-geometrischen Vorstellungen sind somit für die begriffliche Fassung einer Theorie

---

<sup>19</sup> siehe Ref. /7/

autonomer Systeme völlig obsolet. Jede experimentelle Messung reduziert das zu untersuchende Objekt zu einem System mit einem Anfangs- und einem Endzustand und damit zu einem "offenen System", an dem generell nur physikalisch-chemische Parameter gemessen werden können. Kognitive Prozesse lassen sich damit nicht erfassen, d.h. die Thematik von "Kognition und Autonomie" geht auf diesem Weg verloren.

Was benötigt wird, ist ein geeigneter Kalkül zur Entwicklung eines Modells biologisch kognitiver Netzwerke, der eine simultane (formale) Darstellung offener und geschlossener Netze (Systeme) erlaubt. Dafür werden jedoch wenigstens drei unterschiedliche, miteinander jeweils vermittelte Beschreibungspositionen benötigt, aus denen heraus sich das Modell biologisch kognitiver Netze (Systeme)

- als **offene Netzwerke** (Systeme-physikalisch-chemische Beschreibungsdomäne),
- als **geschlossene Netzwerke** (Systeme-kognitive Beschreibungsdomäne), und
- als **Relation von offenen und geschlossenen Netzwerken** (Systemen)

widerspruchsfrei thematisieren lässt. Die wissenschaftliche Umsetzung dieser Forderung stellt primär ein wissenschaftslogisches, interdisziplinär-orientiertes Problem dar, worauf in der Vergangenheit schon mehrfach hingewiesen wurde <sup>20</sup> und begründet gleichzeitig das formale Fundament einer modernen allgemeinen Systemtheorie.

**How to cite:**

E. von Goldammer & J. Paul: Autonomie in Biologie und Technik, in: [www.vordenker.de](http://www.vordenker.de) (Edition: Juni 2002, J. Paul (Ed.)), URL: < <http://www.vordenker.de/autonomie/autonomie.pdf> > — originally published in: Jahrbuch für Komplexität in Natur-, Sozial- und Geisteswissenschaften, Band 6: "Realitäten und Rationalität" (A. Ziemke & R. Kaehr, hrsg.) Duncker & Humblot, Berlin, 1995, S. 277-298.

Copyright 2002 vordenker.de

*This material may be freely copied and reused, provided the author and sources are cited*

**vordenker**

ISSN 1619-9324

<sup>20</sup> siehe Ref. /7/, /10/, /11/, /12/