

Proseminar SS 06: Einführung in die Kognitionswissenschaft – 9. [8.] Sitzung [27.06.06]:
Kognitionswissenschaft und Künstliche Intelligenz

Thesenübersicht zu: Mainzer, K. (1994) *Computer – Neue Flügel des Geistes? Die Evolution computergestützter Technik, Wissenschaft, Kultur und Philosophie*, Berlin / New York + Churchland, P. M. (1992) *Artificial Intelligence + Neuroscience*. In: Churchland, P. M.: *Matter and Consciousness. A Contemporary Introduction to the Philosophy of Mind*, 4. Aufl., Cambridge, MA/London (Engl.), 99–165.

- (1) Zwei große Ansätze zur Entwicklung intelligenter Systeme: (I) *Programmgesteuerte Computer* und (II) *Komplexe dynamische Systeme / Neuronale Netzwerke*
- (2) *Ansatz (I)*: Programmgesteuerte Computer ab den 40er Jahren des 20. Jh. hatten zur Voraussetzung den logisch-mathematischen Begriff einer universellen seriellen Rechenmaschine (Turing (GB)) und deren technisch-physikalische Verwirklichung (Zuse (D), von Neumann (D / USA)).
- (3) Die direkte und vollständige Berechnung (*algorithmische Computation*; vgl. Churchlands (1992, 107-112) Fallbeispiel eines Schachprogramms sowie Newell (1989, 65)) stößt v.a. aufgrund des *Rahmenproblems* (= explosionsartiges Anwachsen der kombinatorischen Optionen oder „Züge“ in einem Suchbaum oder Problemlösungs-Raum (*problem space*)) sehr schnell an unüberwindliche Grenzen. Dies führte im Rahmen des Ansatzes (I) zur Entwicklung der eigentlichen KI: automatische kreative und intelligente Suchverfahren und *Heuristiken* (informelle Faustregeln) für Problemlösungen in Datenstrukturen (vgl. Churchland 1992, 111-112). Das Epochenjahr mit dem Durchbruch dieses Ansatzes war 1956 (J. McCarthy, A. Newell, H. Simon). Die logische Kalküle verkörpernden Programmiersprachen PROLOG und LISP siedeln sich auf dieser Ebene an.
- (4) Universelle KI-Programme beanspruchen faktisch und meist auch ausdrücklich die Verwirklichung von G. W. Leibniz' Vision einer *mathesis universalis* (universelle Wissenschaft), die eine deskriptive mathematische Formelsprache (*characteristica universalis* / formalisierte logische Grammatik), und ein logisch-mathematisches Ableitungssystem (*ars combinatoria* / informationsverarbeitende Algorithmen) auf einer Datenbasis aus Definitionen und analytisch-notwendigen sowie empirisch-kontingenten Sätzen (*ars iudicandi* / abduktiv-induktive Analysis) umfasst.
- (5) Die universellen KI-Programme arbeiten mit *formalen* logisch-mathematischen Datenspeichern und Rechenoperationen, aufbauend auf dem Binärkode der Maschinensprache über imperativische Programmiersprachen (FORTRAN, ALGOL → AL) zu prädikativen Programmiersprachen (z.B. PROLOG → AL + PL 1. Stufe mit Relationen) und schließlich funktionalen Programmiersprachen (z.B. LISP → AL + PL + Funktionen (log: λ -Kalkül).
- (6) Auch diese *universellen* kreativen, heuristischen Kalküle (= *General-Problem-Solver-Programme*) stießen bald (ca. 1962) an ihre Grenzen, die zur Einführung *spezifischer* (inhaltlicher) Regeln und *spezifischer* (inhaltlicher) heuristischer Rahmenbedingungen führten.
- (7) Aus der Situation in (6) entstehen schließlich die nun vorherrschenden lokalen, speziellen *Expertensysteme*, d.h. wissensbasierte Systeme (Rechenprogramme als *intelligent agents*) für bestimmte Gegenstandsbereiche (Pionier: Marvin Minsky). Sie speichern und verarbeiten inhaltliches Wissen in Wissensrepräsentationen oder deklarativen Datenbasen (= symbolische statt nur numerische Informationsverarbeitung: *content theory*) und modellieren in den Rechenoperationen die Urteilskraft und kognitiven Mechanismen eines menschlichen Experten als Aufgaben- bzw. Problemlöser in einem begrenzten Aufgabengebiet (*process model*). Über AL, PL, Relationenlogik und mathematische Funktionen hinaus kommen nun logische und wissenschaftstheoretische Methodologien (kontextsensitive Abduktion, Induktion, Erklärung, Wahrscheinlichkeitswerte, Randbedingungen (*constraints*) etc.) hinzu. Frühe Beispiele für Anwendungsprogramme sind DENDRAL (Chemie) und MYCIN (Medizin), während moderne objektorientierte Programmiersprachen wie C++ standardmäßig über das Instrumentarium zur Bewältigung der hier anstehenden Herausforderungen verfügen.
- (8) Expertensysteme verfügen in der Regel über folgende Binnenstruktur:

- (a) Wissensrepräsentation (Datenbasis) aus deklarativem Fakten- und prozeduralem Regelwissen
 - (b) Ableitungssystem (Problemlösungskomponente)
 - (c) Erklärungskomponente: Das Programm erklärt, wie und warum es zu einer bestimmten Problemlösung gekommen ist.
 - (d) Wissenserwerb: Heuristiken für die Erweiterung der Wissensbasis
 - (e) Dialogkomponente: kommunikative Mensch-Maschine-Schnittstelle.
- (9) Expertensysteme geben den Anspruch auf eine universelle Simulation des Geistes auf und beschränken sich auf lokale Unterstützung des menschlichen Geistes.
- (10) Expertensysteme verfahren in der Frage der Wissensrepräsentation pragmatisch. Je nach Aufgabe und Eignung werden unterschiedliche Ansätze gewählt und auch kombiniert. In der Regel fußt die Wissensbasis auf der Prädikatenlogik (siehe den „Überblick über die Disziplin KW“ im Seminarprogramm, Abschnitt (E), 1 [S. 4]). Das Ableitungssystem ist meist ein regelbasiertes Produktionssystem in der Form von Wenn ... Dann – Regeln (siehe den „Überblick über die Disziplin KW“ im Seminarprogramm, Abschnitt (E), 2 [S. 5]). Dazu gehören auch Regelinterpretationen und Regelverknüpfungen in Vorwärts (*data driven*) und Rückwärtsverkettung (*goal driven*) (vgl. Churchland (1992, 107-112). Verwendet werden auch Begriffsstrukturen, d.h. semantische Netze bzw. empirische kognitive Schemata (*frames* mit *slots* + *fillers*, siehe den „Überblick über die Disziplin KW“ im Seminarprogramm, Abschnitt (E), 3 [S. 5]). In der Programmiersprache C++ entspricht dies den dort zentralen Klassen bzw. Objekten bestehend aus strukturellen Schemata (Konstruktoren), Eigenschaften (Instanzvariablen) und Verhaltensweisen / Operationen (Funktionsklassen oder Methoden). Im weiteren finden in Expertensystemen z.T. sehr ausgeprägt auch Analogien Verwendung (siehe den „Überblick über die Disziplin KW“ im Seminarprogramm, Abschnitt (E), 4 [S. 5-6]) sowie natürlich piktoriale, visuell-räumliche Repräsentationen mit Zooming-, Scanning-, Rotations- und Transformationsoperationen (siehe den „Überblick über die Disziplin KW“ im Seminarprogramm, Abschnitt (E), 5 [S. 6]).
- (11) Die Stärke programmgesteuerter Computer und Expertensysteme ist die Repräsentation und Bearbeitung hochstufigen Wissens und Könnens: Logik – Grammatik – Rechnen – Expertenwissen – wissenschaftliche Erklärungen – Bild- und Musterverarbeitung – Automatisches Beweisen.- Robotik.
- Anmerkung: Das globale Arbeitsmodell der KI und KW von Newell (SOAR) ist die wichtigste programmatische Vereinheitlichung des Ansatzes (I) unter biologischen und psychologischen Realzeitbedingungen. Newell beschreibt sein Modell selbst als „bus-oriented system“, d.h. als nur die zentrale Plattform von Prozessor + RAM (Arbeitsspeicher), Hauptplatine (Datentransfer) und Datenspeicher (Gedächtnis) simulierend. Riesige und zentrale Bereiche der Kognition werden nicht thematisiert, sondern wären, so Newell, als zusätzliche Module an das Bus-System anzuhängen. Nicht von SOAR erfasst sind Begriffserwerb, Spracherwerb und -verarbeitung, Metaphorik und Analogie, Wahrnehmungsanalyse, motorische Planung, kognitive Entwicklung, Onto- und Biogenese, Bewusstsein, neurobiologische Realisierung, Imagination, Motivation, Emotionen, normative Werte.
- (12) Außer der nur lokalen Verwendbarkeit haben Expertensysteme auch innere Grenzen hinsichtlich der (a) kognitiven Subjektivität und emotionalen Intelligenz, (b) der ganzheitlichen intuitiven Einsicht sowie (c) des allgemeinen holistischen Hintergrundwissens (vgl. Churchland (1992, 118-122).
- (13) *Ansatz (II)*: Komplexe dynamische Systeme / Neuronale Netze ist statt von Rechenprogrammen und -maschinen von den informationsverarbeitenden Methoden natürlicher physikalischer und biologischer Systeme inspiriert.
- (14) Nach dem 2. Hauptsatz der Thermodynamik tendieren abgeschlossene Systeme zu maximaler Entropie oder Unordnung, d.h. zum Abbau von Ordnung und Komplexität. Komplexe lebende Systeme, also biologische Ordnungsstrukturen, müssen daher offene Systeme sein, fernab des thermischen Gleichgewichts. Erhalt und Aufbau von Ordnung in der Physik und Biologie hängt erstens von ständiger Zufuhr von Energie von außen ab. Zweitens zeigt die positive empirische Datenbasis der Naturwissenschaften, dass spontane Selbstorganisation (Erzeugung qualitativ neuer Strukturen und Phänomene) in Form von Mikroevolution begrenzter Reichweite eine extreme Leistungsfähigkeit aufweist. Dies scheint jedoch nicht ohne weiteres auf eine Selbstorganisation in makroevolutiver Reichweite ausweitbar, d.h. auf die Hervorbringung ontologisch eigengesetzlicher neuer Ebenen: „Es gibt keinen theoretischen Grund, der erwarten lassen würde, daß evolutive Linien mit der Zeit an Komplexität zunehmen; es gibt auch keine empirischen Belege, daß dies geschieht“ (Szathmáry, E. / Maynard Smith, J. (1995) The major evolutionary transitions. In: *Nature* 374, 227-223). Hier müssen entweder spekulative Faktoren wie bisher unbekannte Mechanismen postuliert werden (so

heute die - meist implizite - Mehrheitsmeinung), oder aber zusätzliche Organisationsprinzipien (Informationsursachen) komplexer Systeme angenommen werden (Aristoteles, Leibniz, Kant, Heisenberg, Dürr, Barrow, Kutschera, in etwa auch Hawking).

- (15) Ein dynamisches System ist ganz allgemein eine strukturierte Menge von Elementen mit internen + externen Relationen materieller, energetischer und informationeller Flüsse. Ein System der klassischen deterministischen Physik wie auch ein einfaches probabilistisches System der Chaostheorie lässt sich analytisch-geometrisch durch 3 Ortskoordinaten und eine Kraftkomponente darstellen (Hamilton-System): als Zustand / Phase zum Zeitpunkt t plus der Menge Z der möglichen Phasen (= Phasenraum). Die Zustandsentwicklung eines Systems ist seine Bahnlinie (Trajektorie) im Phasenraum. Qualitative Phasenübergänge in offenen Systemen können zu hierarchisch organisierten, emergenten Strukturen führen. Das heißt: Mikroskopische lokale Wechselwirkungen eines komplexen Systems erzeugen in Phasenübergängen makroskopische Ordnungsstrukturen. Einfache Systeme führen relativ unabhängig vom Anfangszustand zu einem konstanten oder periodischen Endzustand. Komplexe Systeme sind extrem abhängig vom Anfangszustand, gehorchen also einer nichtlinearen Dynamik, und führen zu lokalen und globalen Gleichgewichtszuständen oder Attraktoren als Endzuständen (Fixpunkte, geschlossene Kurven, beschränkte Flächen, Grenzzyklen und seltsame fraktale, d.h. iterative Attraktoren).
- (16) Die Entwicklung von Ordnungsstrukturen in dynamischen Systemen kann durch sog. Zelluläre Automaten (die ursprüngliche Idee stammt von v. Neumann) simuliert werden, bestehend aus einer Zustandsmenge Z , einem Grundalgorithmus, einer Eingabemenge mit Eingangsvektoren $X = (x_1, x_2, \dots, x_n)$ und einer Ausgabemenge mit Ausgabevektoren $Y = (y_1, y_2, \dots, y_n)$. Man unterscheidet Automaten der Klasse 1 mit einfachen Attraktoren (1 Fixpunkt), der Klasse 2 mit Grenzzyklen (Oszillationen), der Klasse 3 mit seltsamen Attraktoren (chaotische, fraktale Muster), der Klasse 4 mit komplexen Strukturen (nicht vorhersagbares Entstehen neuer Ordnungsstrukturen). (Anm: Zelluläre Automaten bewältigen auch die in der Physik zentralen Differentialgleichungen und vermögen universelle serielle Rechenmaschinen (Standardcomputer) zu simulieren).
- (17) Einfache künstliche neuronale Netzwerke (1-schichtige Hopfield-Systeme) sind parallel verteilte, robuste und fehlertolerante dynamische Systeme. Sie sind nach einer Lernphase zu eintrainierter Mustererkennung in der Lage, die mathematisch in Phasenübergängen zu Attraktoren (Gleichgewichtszuständen) besteht.
- (18) Anspruchsvollere 3-schichtige Netzwerkhierarchien sind nach entsprechender Eintrainierung zur spontanen Klassifizierung von Merkmalen und zielgeleiteten motorischen Aktionen nach Grundsätzen der Selbstorganisation in der Lage. Die Repräsentation (Kodierung) der sensorischen und motorischen Muster geschieht in mehrdimensionalen Phasenräumen, wobei der jeweilige Zustand durch einen n -dimensionalen Kodierungsvektor (Matrix) dargestellt wird. Die intersensorische und sensomotorische Koordination erfolgt durch Vektortransformationen (Matrizenrechnung). Der Trainings- und Lernprozess des Systems fußt auf der Tatsache, dass das innere Produkt zweier Vektoren eine Maßzahl für die Ähnlichkeit zweier Vektoren ist. Dies erlaubt, in 1000- und 10000-fach wiederholten Versuchen die Einzelwerte des Ist-Vektors mit dem angestrebten prototypischen Soll-Vektor abzugleichen und immer besser anzupassen (z.B. im Back-Propagation-Algorithmus).
- (19) Hierarchische neuronale Netze bilden interne Modelle der Außenwelt nach entsprechendem Training und können mittels des Dateneingangs weiterlernen. Sie identifizieren Wahrnehmungsmerkmale, tieferliegende strukturelle Kategorien und Sprachmuster (vgl. Churchland 1992, 146-155).
- (20) Die Anfälligkeit hierarchischer Netze für fehlerhafte nur lokale Minima (Endzustände) wird durch sog. probabilistische Netzwerke verbessert (Boltzmann-Maschinen), die nach einem überwachten Lernprozess globale Minima generieren (Hinton, Sejnowski). Noch weiter gehen unterschiedliche Lernalgorithmen für selbstständige Lernprozesse (z.B. spontane Begriffsbildung) in Zustands- oder Vektorräumen, besonders wenn sie einer 3-dimensionalen neuronalen Geometrie / Topographie der Außenwelt + n -dimensionalen motorischen Bezugssystemen zugeordnet sind (Kohonen, Gärdenfors). Vgl. die topographischen somatosensorischen und motorischen Körperkarten des realen Gehirns: Die *Natur* qua Natur arbeitet wie die Neuroinformatik v.a. mit analogen, geometrisch-topographischen Repräsentationen, die sprachlich-mathematische *Kultur* arbeitet wie die traditionelle Informatik und KI v.a. mit algebraisch-analytischen Repräsentationen.
- (21) Bisher nicht gelöste Hypothesen des Ansatzes (II) sind die extrem lange überwachte oder auch selbstständige *Trainingsphase*, die in der Regel eine zusätzliche sorgfältige Aufbereitung des Daten-

materials erfordert. Sie steigt angesichts der tatsächlichen, extrem komplexen und anspruchsvollen menschlichen Wahrnehmungs- und Kognitionsleistungen rapide ins Unermessliche. Dazu kommt die von vorne herein in Form von komplizierten *Algorithmen* in die künstlichen neuronalen Netze einzubauende „künstliche“ Intelligenz: „Much of the interesting work has been done in setting up the innate wiring of the network“ (Pinker 2002, 83). Bei der menschlichen Kognition erscheint diese auch hier voraussetzende apriorische Intelligenz nur durch genetische, angeborene Strukturen der Wissensrepräsentation und -verarbeitung erklärbar, nicht allein aus aktuellen selbstorganisierenden Prozessen (Pinker 2002).

Mainzers Fazit ist daher zum einen: „Die wesentlichen Lern- und Selbstorganisationsstrukturen des lebenden Gehirns [sind] noch nicht bekannt“ (724). Zum anderen argumentiert Mainzer dafür, dass semantische Bedeutungen, Intentionalität und Subjektivität irreduzible, nicht identisch reproduzierbare Eigenschaften neuronaler Netzwerke als offener dynamischer Systeme sind (779): „Die intentionalen Symbolsysteme der menschlichen Kultur“ ermöglichen einen „kulturell-geistesgeschichtlichen Sozialisationsprozeß“, der „Einmaligkeit“ und „Würde“ des Menschen ausmacht (803).

- (22) Der Ansatz (II) bewältigte ursprünglich und besonders gut niedrigstufiges Wissen wie Signalerkennung und -analyse, Bilderkennung und -analyse, Sprachanalyse, Wahrnehmungs- und Bewegungsanalyse. Insbesondere erlaubt der Ansatz (II) eine elegante Handhabung der in der menschlichen Kognition kaum zu überschätzenden Analogie und Metaphorik (siehe den „Überblick über die Disziplin KW“ im Seminarprogramm, Abschnitt (E), 4 [S. 5]). Kontrovers diskutiert wird zur Zeit, ob der Ansatz (II) nicht auch das gegenüber dem Ansatz (I) leistungsfähigere Modell für hochstufige Kognition ist. Bejaht wird dies u.a. von Churchland, Metzinger und Mainzer, verneint von Fodor, Pylyshyn und Pinker.
- (23) Die Synergetik (Haken) ist ein konzeptgeleiteter (*top-down*) Ansatz der Theorie komplexer dynamischer Systeme und neuronaler Netzwerke. Er untersucht die allgemeinen Gesetze von Selbstorganisation: Anfangsfluktuationen – Nichtlinearität – Symmetriebrechungen – Ordnungsstrukturen.
- (24) Die *angewandte* KI umfasst heute u.a.
- (a) computergestützte Mathematik, insbesondere bei – große Rechenkapazität erfordernden – *Numerischen Näherungsverfahren* mit vielen kleinen diskreten Intervallen bzw. Rechenschritten bei der mathematischen Modellierung und Berechnung physikalischer, technischer und wirtschaftlicher Sachverhalte und Vorgänge. Diese sind immer dann notwendig, wenn die klassische Analysis bzw. Differentialgleichungen mit ihren stetigen und in Funktionen komprimierbaren Veränderungen der Größen versagen. Zu (a) gehören ferner *Mathematisches Beweisen*, *Computer aided design* (CAD) und *Numerische Rekursion* in der fraktalen Geometrie, die in der realen Welt vorherrscht.
 - (b) computergestützte Naturwissenschaft: Ab drei Objekten sind physikalische Systeme rechnerisch irreduzibel, d.h. nicht mehr deterministisch und mechanisch berechenbar, sondern nur noch in chaostheoretischer Näherung und Wahrscheinlichkeit: „Für die moderne Astronomie und Raumfahrt ist die numerische Approximation und Simulation von Mehrkörperproblemen grundlegend“ (470). Und: „Die philosophische Quintessenz der Chaostheorie lautet also: Weder die physikalische Welt, noch unsere alltägliche Lebenswelt ist ein programmgesteuerter Digitalcomputer, in dem jeder Schritt von uns plan- und prognostizierbar wäre“, wir sehen uns im Gegenteil einem „Reich der Freiheit“ gegenüber (466). Zur angewandten KI gehört auch molekulares Design und Modellierung in der Chemie, mit der dort typischen Verbindung numerischer Daten mit qualitativer, symbolischer und geometrisch-imaginaler Information in den chemischen Strukturformeln; außerdem die Analyse nichtlinearer Prozesse in der medizinischen Kardiologie, Neurophysiologie und Psychopharmakologie.
 - (c) Wissenschaft und Wirtschaft: Analyse von Forschungsparadigmen und ökonomischen Entwicklungen als Attraktoren in dynamischen Forschungs- bzw. Wirtschaftsprozessen.
 - (d) Medieninformatik und interaktive virtuelle Welten, z.B. 3-dimensionale Animation mit Datenbrille + Datenhandschuh.
 - (e) spekulative Szenarios wie Quantencomputertechnologie und Biochips aus Kettenmolekülen der organischen Chemie mit einer Speicherkapazität von 1 Million Festplatten à 120 GB auf 1 cm Länge. Bioadapter (Dyson, Tipler), d.h. Direktverbindungen zwischen Gehirn und Computerprogrammen bzw. Personoiden als Software-Duplikaten von Menschen scheinen dagegen aus grundsätzlichen mathematischen und physikalischen Erwägungen nicht durchführbar zu sein (vgl. Barrow 1992).