

Zusammenfassung Referat

„Was ist Cognitive Science?“

Seminar SS 08, Dr. Karl Leidlmair

Oberhauser Alexander

Schöndorfer Marion

Windisch Susanne

Was ist Cognitive Science?

Ich befürchte, wir werden keine genaue Antwort auf diese Antwort finden.

- 1. Was wir herausfinden können sind Definitionen!
- 2. Wir können versuchen kognitive Fähigkeiten des Menschen zu beschreiben!
- 3. Oder wir können schauen was Menschen machen die sich als Cognitive Scientists bezeichnen!

1. Definition

Cognitive Science ist die Wissenschaft, die versucht, zu verstehen, wie unser Gehirn funktioniert. Kognitive Fähigkeiten wie Wahrnehmung, Denken, Lernen, Motorik und Sprache werden untersucht und modelliert.

Cognitive Science ist interdisziplinär: Eine zentrale Rolle spielen Informatik, Mathematik, Neurowissenschaft, Psychologie, Linguistik und Philosophie.

Die Cognitive Science ist eine theoretische Disziplin. Im Unterschied zu ihren beiden Nachbardisziplinen, der Kognitionspsychologie, die kognitive Prozesse unter empirischen Bedingungen untersucht, und der Künstlichen-Intelligenz-Forschung als einer technischen Disziplin geht es der Cognitive Science um eine rein abstrakte Beschreibung von Kognition.

Kognitive Prozesse seien, so meint jedenfalls die Cognitive Science, gleichzusetzen mit *formalen* Beschreibungen von kognitiven Prozessen.

Kognitive Prozesse seien daher auch unabhängig von ihrer jeweiligen physikalischen Realisierung erfassbar. Aus dieser Annahme folgt auch, dass Intelligenz auf beliebigen Hardwareträgern sich realisieren lassen müsste.

Im Extremfall müsste es möglich sein durch eine detaillierte formale Beschreibung der Gedanken eines Menschen diese Gedanken so „abschöpfen“ zu können, dass, selbst wenn sein Körper eines Tages nicht mehr vorhanden ist, seine Gedanken in einem Computer weiterleben. Denn was diese Gedanken essentiell ausmache, wäre ja nicht die „Neuronenhardware“ auf der sie rein per Zufall ablaufen, sondern ihre formale Beschreibung.

2. Die CS vermittelt eine Ahnung davon, wie unser Gehirn funktioniert. Also den Kognitiven Fähigkeiten des Menschen.

Das menschliche Gehirn ist das komplexeste System des uns bekannten Universums. Um dieses System auch nur ein kleines bisschen zu verstehen, genügt es nicht, sich aus nur einer Richtung anzunähern.

Menschen verfügen über zahlreiche verschiedene kognitive Fähigkeiten: Gedächtnis, Sprache, Wahrnehmung, Problemlösen, geistiger Wille, Aufmerksamkeit usw. Ziel der Kognitionspsychologie ist es, die Eigenarten dieser Fähigkeiten zu erforschen und, soweit wie möglich, in formalen Modellen zu beschreiben. Diese Modelle können dann als *kognitive Architektur* auf einem Computer realisiert werden. Auch die künstliche Intelligenz (KI) hat das Ziel, kognitive Fähigkeiten in Maschinen zu realisieren. Dabei dürfen die künstlichen Agenten jedoch – im Gegensatz zu kognitiven Architekturen – auch Strategien verwenden, die von Menschen nicht genutzt werden

1. Grundlagen: Proteine, Enzyme, Aufbau von Zellen
2. Eigenschaften und Aufbau von Nervenzellen
3. Intrazellulärer Transport von Stoffen
4. Die Struktur der Zellmembran
5. Wie elektrische Spannung in Nervenzellen entsteht (Ruhepotential)
6. Wie Signale entstehen und weitergeleitet werden (Aktionspotential)
7. Wovon die Weiterleitung von Signalen abhängt
8. Wie Synapsen funktionieren
9. Wie Gifte und Drogen wirken
10. Wie wir lernen
11. Wozu Gliazellen gut sind (von denen es 10x mehr als Nervenzellen gibt)
12. Warum wir manche Dinge kaum kontrollieren können (Herzschlag, Reflexe)

Linguistik - Sprachwissenschaft

Linguistik ist aus zweierlei Gründen wichtig für die Kognitionswissenschaft. Einer davon, die Sprachbeherrschung gehört zu den herausragenden kognitiven Fähigkeiten, eine angeborene kognitive Fähigkeit, des Menschen— und nur des Menschen. Das Verfügen über Sprache ist zudem Voraussetzung für das Verfügen über einige andere kognitive Fähigkeiten. Ohne Sprache könnten zumindest viele Gedanken nicht gedacht und viele

Probleme nicht gelöst werden. Deshalb muss jede Theorie des menschlichen Geistes auch das Phänomen Sprache beachten.

Der zweite Grund: Linguistik ist Voraussetzung für die Computerlinguistik, die wiederum ein Teilgebiet der künstlichen Intelligenz ist. Hier wird versucht, Aufgaben zu automatisieren, die ein gewisses "Verständnis" von Texten erfordern (maschinelle Übersetzung, Textzusammenfassung, etc.).

In der Kognitionswissenschaft hat die Sprache daher immer eine zentrale Rolle gespielt. Zum einen stellt sich die Frage, wie die menschliche Sprachbeherrschung möglich ist, zum anderen, wie man Maschinen zur Sprachbeherrschung bringen kann.

3. Cognitive Scientists – Aufgabenbereiche und Wissensgebiete

- Cognitive Scientists arbeiten und forschen im Bereich der künstlichen Intelligenz. Studienrichtungen wie Mathematik oder Informatik sind dabei eingeschlossen aber sind alleine zu allgemein (zu wenig Fokus auf das Verständnis von Intelligenz) und zu spezifisch (Herangehensweise auf ein Wissenschaftsgebiet beschränkt).

Ethik der Kognitionswissenschaft

In Japan werden Roboter dafür eingesetzt, für allein stehende alte Menschen zu sorgen! Die ethische Bedeutung von oberflächlich menschenähnlichen Robotern ist keineswegs klar, geschweige denn die von Maschinen mit "wirklicher" Intelligenz.

Es gibt wenige Wege, in der CogSci auf denen man nicht mit ethischen Fragestellungen konfrontiert wird.

3. Geschichte der Cognitive Science

Kognitive Wende

Die Entwicklung der Kognitionswissenschaft hängt eng mit der so genannten „kognitiven Wende“ zusammen, deren Höhepunkt auf das Jahr 1956 fällt. Bis dahin hatte in der Psychologie und der Philosophie des

Geistes der Behaviorismus eine maßgebliche Rolle gespielt, der sich bei der Beschreibung intelligenten Verhaltens ausschließlich am Schema Reiz und Reaktion orientiert und dabei innere Verarbeitungsprozesse ausklammern möchte. Vom Behaviorismus hin zum Kognitivismus.

Als Reaktion auf eine derart einseitige Sichtweise geht es in der kognitiven Wende um eine Wiedereinführung der Sprache des Geistes in den wissenschaftlichen Diskurs. Es geht um die Suche nach einem geeigneten Modell für die Beschreibung der inneren Verarbeitungsprozesse im Kopf.

Als besonders brauchbar hat sich hierbei die so genannte komputationale Theorie des Geistes erwiesen. Damit ist die These gemeint, dass das Gehirn ein informationsverarbeitendes System sei und prinzipiell wie ein Computer arbeite. Die Unterscheidung zwischen Geist und Gehirn lasse sich analog zu der Unterscheidung zwischen Software und Hardware verstehen. So wie die Software durch Datenstrukturen und Algorithmen bestimmt sei, sei der Geist durch mentale Repräsentationen und Rechenprozesse bestimmt. So wie die abstrakte Beschreibung der Software möglich ist, ohne direkt die Hardware zu untersuchen, sollte eine abstrakte Beschreibung der geistigen Fähigkeiten möglich sein, ohne direkt das Gehirn zu untersuchen. Und so, wie die Existenz einer Softwareebene problemlos mit dem Materialismus zu vereinbaren ist, sollte auch die mentale Ebene in eine materialistische Interpretation eingebettet sein. Das Verhältnis von kognitiven Prozessen zu ihren physikalischen Trägerprozessen wird dabei in Analogie zum Verhältnis von Software und Hardware eines Computers bestimmt.

Bei dieser notwendigen Korrektur des Behaviorismus darf man allerdings nicht übersehen, dass letzterer seinerseits aus einem Erklärungsnotstand einer ganz bestimmten Denkströmung Anfang des 19. JHs hervorgegangen ist, nämlich des so genannten alten Mentalismus der Würzburger Schule in der Psychologie.

Der alte Mentalismus

Die Würzburger Schule wurde um 1900 von Oswald Külpe begründet. Eine seiner Grundannahmen war, dass psychische Phänomene qualitativ verschieden sind von physikalischen Phänomenen und daher nicht auf diese reduziert werden können. Die eigene Erfahrung in der Form der systematischen Selbstbeobachtung wird in der Lehre der Würzburger Schule als Grundquelle psychologischer Erkenntnisse angesehen. Wesentliche Methode, da psychische Phänomene nicht übersetzbar sind in das von außen beobachtbare Verhalten, war die experimentelle Erfassung kognitiver Prozesse mit Hilfe der Retrospektion (rückschauende Introspektion).

Das Problem dieses Ansatzes war freilich, dass er keine Objektivität für sich in Anspruch nehmen konnte.

Der Behaviorismus

Der Behaviorismus war ein Versuch wissenschaftlich objektive Methoden in der Psychologie zu etablieren, er war als eine Reaktion auf die Probleme der Introspektion als einer psychologischen Forschungsmethode entstanden. Introspektive Berichte über das mentale Innenleben waren für die Wissenschaftler nicht von außen überprüfbar. Der Behaviorismus zog daraus die Konsequenz, dass sich die Psychologie auf eine Erforschung des Verhaltens beschränken müsse.

Der **Behaviorismus** ist ein wissenschaftstheoretischer Standpunkt, der zugrunde legt, dass das Verhalten von Menschen und Tieren mit den Methoden der Naturwissenschaft untersucht werden kann. Er versteht sich somit als eine *Theorie der Wissenschaft vom Verhalten*, der Verhaltenswissenschaft oder Verhaltensanalyse.

Grundsätzlich geht es darum menschliches Verhalten zu erklären im Schema von Reiz und Reaktion.

Der Behaviorismus, so wird jedenfalls von Vertretern der so genannte Folk psychology argumentiert, stehe im Widerspruch zu unserer Alltagspsychologie. In der Alltagspsychologie verwenden wir ständig psychologische Zuschreibungen, um das Verhalten eines Menschen erklären zu können. Wir unterstellen unserem Gegenüber intentionale Einstellungen mit dem Ziel, daraus auf sein Verhalten schließen zu können. Wenn wir uns an die Alltagspsychologie halten dann dürfen wir nicht wie dies der Behaviorismus tut psychische Phänomene als Ursachen für physikalische Phänomene von vornherein ausschalten. Der Behaviorismus wollte die Psychologie neu begründen als eine Naturwissenschaft, der Preis ist ganz einfach dass er unser Verhalten nicht erklären kann.

Aktuelle Entwicklungen

Das Computermodell des Geistes ist in den letzten Jahrzehnten einer scharfen Kritik unterzogen worden. Diese Kritik hat im Wesentlichen zwei Quellen:

- Zum einen hat sich die Beschreibung des Gehirns durch die Kognitive Neurowissenschaft rasant entwickelt. Dies zeigt sich etwa in der zunehmenden Bedeutung von bildgebenden Verfahren, die es unplausibel machen, das Gehirn bei der Erforschung des Geistes nicht zu beachten.
- Zum anderen haben sich andere erfolgreiche Ansätze entwickelt, so z. B. der Konnektionismus und die Modellierung von neuronalen Netzen.

- Andere alternative Paradigmen in der Kognitionswissenschaft sind z. B. der Dynamizismus, Künstliches Leben (Artificial Life) und die verkörperlichte und situierte Kognitionswissenschaft.

„Künstliches Leben“ ist ein Terminus, der der Künstlichen Intelligenz gegenübersteht: anstatt abstrakte Aufgaben zu lösen (wie z. B. Schachstellungen analysieren), was uns Menschen oft schon wegen der bloßen Anzahl der Lösungsmöglichkeiten schwierig erscheint, Computern jedoch leicht fällt, solle man erst die Bewältigung der vermeintlich profanen Alltagsprobleme verstehen. Viele Aufgaben, die uns einfach erscheinen (wie z. B. Laufen, Freunde und Feinde erkennen, Wunden heilen, einen Ball fangen ...) sind von Computern oder Robotern derzeit noch gar nicht oder nur sehr eingeschränkt zu bewältigen.

Die verkörperlichte und situierte Kognitionswissenschaft wiederum geht davon aus, dass Kognition nicht ohne Bezug auf einen spezifischen Körper (Verkörperlichung, Embodiment) und eine spezifische Umgebung (Situiertheit) erklärt werden kann. Diese Forderungen resultieren aus dem Zweifel daran, dass Kognition ein Prozess ist, der sich in einer Welt der abstrakten symbolischen Repräsentationen vollzieht, verhältnismäßig unabhängig von der genauen Sensorik, Motorik und dem zeitlichen Geschehen in der Außenwelt.

Die Kritik am Computermodell des Geistes führte zeitweise zur generellen Infragestellung der Kognitionswissenschaft. Mittlerweile haben sich die Wogen jedoch weitgehend geglättet. Kognitionswissenschaftler verwenden nun selbst auch neuronale Netze und stehen in einem engen Kontakt mit der Kognitiven Neurowissenschaft.

Das Mehrebenen Modell der Intelligenz

Um das Verhalten eines Systems angemessen verstehen zu können, müssen wir dessen innere Zustände kennen.

In den Naturwissenschaften gehen wir von einem physikalisch kausal geschlossenen Weltbild aus. wir haben Eine ganze Kette von Ursache und Wirkung. Die Idee im naturwissenschaftlichen Weltbild warum es auch physikalisch kausal geschlossen ist, dass in dieser ganzen Kette von Ereignissen immer nur physikalische Ursachen vorkommen dürfen es darf nicht sein dass in dieser kette ein Glied ist das physilkalisch nicht erklärbar ist

Auf diese Weise können wir ganz neue Befehle in einer Programmiersprache erzeugen, ja wir können in einer Programmiersprache sogar eine andere Programmiersprache programmieren. Auf diese Weise entstehen verschiedene Schichten der Programmierung

Dieses Prinzip erklärt, wie es möglich ist, in einer Maschinensprache eine höhere Programmiersprache zu erzeugen und in dieser höheren Programmiersprache wiederum ein Programm zu schreiben, das Eingaben und Ausgaben enthält. Höhere Schichten der Programmierung (denken Sie nur an das uns allen bekannte SPSS) sind auf diese Weise in tieferen Schichten (Assemblersprachen, Maschinensprachen usw.) realisiert.

Wichtig dabei ist, dass höhere Schichten der Programmierung in tiefere Schichten *übersetzbar* sind. Höhere Schichten der Programmierung sind also nichts eigenständiges, sie werden nur aus Gründen der Bequemlichkeit und besseren Übersichtlichkeit verwendet. Höhere Schichten der Programmierung sind daher nichts anderes als abstraktere Beschreibungsebenen der tieferen Schichten.

Es ist die Übersetzbarkeit der höheren Beschreibungsebenen in tiefere Beschreibungsebenen, wodurch sich die Cognitive Science vom Leib-Seele-Dualismus unterscheidet.

Erst dadurch, dass sich höhere Schichten unabhängig von ihrer konkreten Realisierung in tiefere Schichten beschreiben lassen, werden kognitive Zustände als relativ eigenständiger Forschungsbereich thematisierbar.

Dieses Schichtenmodell wird von der Cognitive Science nun auf kognitive Prozesse angewandt, u.zw. auf folgende Art und Weise. Die Cognitive Science unterscheidet drei Schichten:

Die Ebene der intentionalen Einstellungen (die semantische Ebene)

Die Ebene der formalen Beschreibungen (die syntaktische Ebene) und schließlich

Die physikalische Ebene.

Besondere Bedeutung kommt in diesem Modell der syntaktischen Ebene zu. Die Syntax hat nämlich in der Cognitive Science eine Art Scharnierfunktion. Ohne die Zwischenschicht der Syntax kann die intentionale Beschreibungsebene nämlich nicht auf die physikalischen Trägerprozesse übertragen werden. Die Semantik wird zuerst formalisiert und die formalisierte Semantik mechanisiert. Die Formalisierung ist also die notwendige Vorbedingung und Vorstufe zur Mechanisierung.

Übersetzbarkeit

CS übersetzt zuerst die Ebene der intentionalen Einstellungen in die syntaktische Ebene, das heißt sie formalisiert zunächst die inhaltliche Ebene unseres Denkens und übersetzt dann erst in einem zweiten Schritt diese formalisierten Gedanken in physikalische Prozesse. Das eigentliche Problem hierbei ist die Übersetzbarkeit von Semantik in Syntax. Denn wenn die Semantik einmal formalisiert wurde, das heißt in eine Art „Software des Gehirns“ umgewandelt wurde, ist es kein Problem mehr, diese Software dann auch auf einer entsprechenden Hardware zu implementieren.

Semantik ist die Lehre von den Bedeutungen sprachlicher Zeichen. Was ist aber unter der Bedeutung eines sprachlichen Zeichens zu verstehen? Nehmen wir ein beliebiges Wort, z.B. das Wort „Tisch“.

Um zu wissen wovon das Wort „Tisch“ handelt, müssen wir seine Bedeutung kennen. Und das Wort „Tisch“ handelt eben von Tischen. Dies ist sein Begriffsumfang bzw. seine Extension. Die Bedeutung ist aber nicht mit der Extension zu verwechseln.

Die Bedeutung eines Wortes ist jenes Vehikel, wodurch wir uns mit dem Wort auf einen ganz bestimmten Gegenstandsbereich beziehen.

Multiple Instanziierung

Nun zum zweiten Punkt: Warum sind höhere Schichten relativ unabhängig von den niederen Schichten? Wir haben doch soeben unter dem Punkt „Übersetzbarkeit“ festgestellt, dass letzten Endes alle höheren Schichten durch die zugrundeliegende physikalische Schicht determiniert sind. Das heißt doch, dass letzten Endes alles auf die materielle Ebene reduziert werden kann. Was soll dann aber noch die relative Eigenständigkeit höherer Schichten bedeuten?

Die Antwort darauf lautet: Betrachten wir das Verhältnis der Schichten von unten nach oben, so sind Ereignisse auf einer höheren Beschreibungsebene durch Ereignisse auf einer niederen Ebene zwar genau festgelegt (ein Ereignis auf einer niederen Ebene ist *genau einem* Ereignis auf einer höheren Ebene zuordenbar), das umgekehrte ist jedoch nicht der Fall: Ein Ereignis auf einer höheren Ebene kann durch verschiedene Ereignisse auf einer tieferen Ebene realisiert werden. Dies ist das Prinzip der multiplen Instanziierung. Man kann sich das Verhältnis der Schichten zueinander grafisch auch wie folgt vorstellen. Erlaubt ist das folgende (die runden Kreise symbolisieren Ereignisse):

Verboten ist dahingegen (da von unten nach oben betrachtet immer nur je ein Pfeil zu einem Ereignis führt):

Bei der Beschreibung eines Ereignisses auf einer höheren Ebene spielt es daher im Grunde genommen auch gar keine Rolle, wie dieses Ereignis auf einer tieferen Ebene realisiert ist.

Die Sprache formaler Systeme

Was ist aber unter einem formalen System zu verstehen? Es handelt sich hierbei um eine Sprache, die aus vorgegebenen Zeichenketten nach bestimmten Regeln weitere Zeichenketten erzeugt. Die Produktionsregeln zur Erzeugung neuer Zeichenketten bedienen sich ausschließlich der äußeren Gestalt der Zeichenketten ohne Rücksicht darauf, was diese bedeuten könnten (falls sie überhaupt etwas bedeuten).

Damit in diesem Zusammenhang überhaupt von einer Sprache die Rede sein kann, benötigen wir dreierlei: Erstens ein Alphabet, also einen ganz bestimmten Zeichenvorrat, zweitens eine Grammatik, die uns sagt, wie wir die Buchstaben des Alphabets zu sinnvollen Wörtern und Sätzen zusammensetzen können und schließlich und drittens einer Entscheidungshilfe, die uns sagt, welche der grammatikalisch sinnvoll gebildeten Sätze in dem System auch gültige Sätze sind.

Was ist eine strukturelle Abbildung?

Der Ausdruck „Code“ kommt aus der Nachrichtentechnik und meint eine Verschlüsselungsvorschrift für die Übertragung von Informationen. Man kann eine solche Vorschrift im weitesten Sinne des Wortes als eine Abbildung interpretieren

Bei einem strukturellen Abbildungsbegriff besteht zwischen den Zeichen und der zu vermittelnden Information lediglich eine rein per Konvention festgelegte Zuordnung. Aber selbst diese Einschränkung erklärt noch immer nicht hinreichend, was unter einer strukturellen Abbildung zu verstehen ist. Eine strukturelle Abbildung hat nämlich noch die folgende wichtige Eigenschaft: dass nämlich der materielle Informationsträger beliebig austauschbar ist.

Transitivität

Die wohl bekanntesten strukturellen Abbildungen sind der binäre Code und der dekadische Code, welche aber nur zwei Beispiele von strukturellen Abbildungen sind. Neben diesen Codes gibt es allerdings noch zahlreiche andere Möglichkeiten zu codieren, wie zum Beispiel den Hexadezimalen Code mit 16 verschiedenen Einstellungsmöglichkeiten. Transitivität bedeutet in diesem Zusammenhang, dass man einen Code in eine andere strukturelle Abbildung übersetzen kann. Bsp.: Eine Person A verschickt eine Nachricht im hexadezimalen Code an eine Person B. Nachdem Person B diese Nachricht dechiffriert hat schickt sie die Nachricht im binären Code an eine Person C. Aufgrund der Transitivität erhält die Person C die gleiche Nachricht wie bei einer direkten

Übermittlung der Nachricht von A nach C im binären Code. Im Prinzip ist es also egal welchen Code wir zum Verschlüsseln der Information verwenden.

Als Standard setzte sich der binäre Code durch, da mit diesem universellen Code jegliche Art von Information, sei es Bild oder Ton, in der digitalisierten Welt ausgedrückt werden kann.

Gültigkeit von Sätzen

Man hatte sich also auf ein Alphabet einer formalen Sprache geeinigt. Ein Alphabet allein macht aber noch keine Sprache aus. Was benötigt man dazu?

Einerseits braucht man eine Grammatik, damit man die Buchstaben zu sinnvollen Aussagen zusammensetzen kann und andererseits das Wissen über die Gültigkeit der Aussage. Es geht also darum, dass semantische Beziehungen durch richtige Aussagen im formalen System wiedergespiegelt werden.

Folgend wird die Gültigkeit von Sätzen mit Hilfe zweier formaler Systeme erklärt, dem MIU-System und dem pg-System.

Das MIU-System

Das MIU-System ist ein Spiel, es beginnt mit einer Zeichenkette und durch 4 verschiedene Regeln können weitere Zeichenketten hinzugefügt werden. Die Zeichen bei diesem Spiel sind nicht interpretiert, sie haben also keine Bedeutung. Daher können die gebildeten Sätze des Systems weder wahr noch falsch sein.

Am Start besitzen wir die Zeichenkette MI. Mit den Regeln sollen nun weitere Zeichenketten zugefügt werden.

Regel 1: Endet eine Zeichenkette mit einem I, so kann ein U angefügt werden.

Regel 2: Bei einer Kette von Mx, können wir eine weitere Kette Mxx erzeugen.

X ist eine beliebige Zeichenkette

Regel 3: Kommt in einer Kette III vor, so kann dies durch ein U ersetzt werden.

Regel 4: Kommt die Kette UU vor, so kann man sie streichen.

Aus dem Startkapital lassen sich also mit Hilfe der Regeln verschiedene Formen von Ketten bilden. Das Startkapital (MI) wird als Axiom bezeichnet,

was einen von vornherein gültigen Satz in einem formalen System beschreibt. Die Spielregeln sind Ableitungsregeln im formalen System. Nun zur Gültigkeit der Sätze. Alle Zeichenketten, die vom Axiom unter Nutzung der Ableitungsregeln erstellt worden sind, sind gültige Sätze in diesem formalen System. Die Gültigkeit bezieht sich allerdings nur auf das formale System und nicht auf das Semantische System, da den Zeichen keine Bedeutung beigemessen wurde.

Das pg-System

Ähnlich dem MIU-System besteht das pg-System allerdings aus einer unendlichen Anzahl an Axiomen und nur einer Ableitungsregel.

Die Axiome lassen sich mit Hilfe einer Schablone erzeugen.

Xp-gx- sei eine gültige Zeichenkette, wenn x immer die gleiche Anzahl an Bindestrichen enthält. Setzt man für x: -- ein, so erhält man das Axiom:

--p-g---

Die Regel: x, y und z seien eine Folge von Bindestrichen. Angenommen xpygz sei ein gültiger Satz in dem formalen System, so sei auch xpy-gz- ein gültiger Satz.

Dabei besteht Variable x aus 2 Bindestrichen, y aus einem und z besteht aus 3 Bindestrichen. Bei Anwendung der Regel erhält man also den folgend gültigen Satz:

--p--g----

Dies erscheint vielleicht ein bisschen verwirrend und unverständlich, was sich aber mit der Interpretation der Zeichenkette bessern sollte. Die Zahl der Bindestriche entspricht hier einer natürlichen Zahl ($-=1$), p bedeutet plus und g bedeutet gleich ($2+2=4$).

Was wir jetzt erkennen ist ein formales System, welches die Addition formalisiert.

Betrachten wir nun die Gültigkeit des Satzes im Sinne des Formalistenmottos: Syntax mirrors Symantics, so erhalten wir nicht nur eine gültige Aussage im formalen System (MIU), sondern wir erhalten auch eine wahre Aussage, denn die syntaktische Beziehung spiegelt die semantische Beziehung wieder. $2+2=4$ ist eine wahre Aussage in der Mathematik.

Wir sehen also den Unterschied zwischen formaler Gültigkeit und semantischer Gültigkeit.

Von der Formalisierung zur Mechanisierung

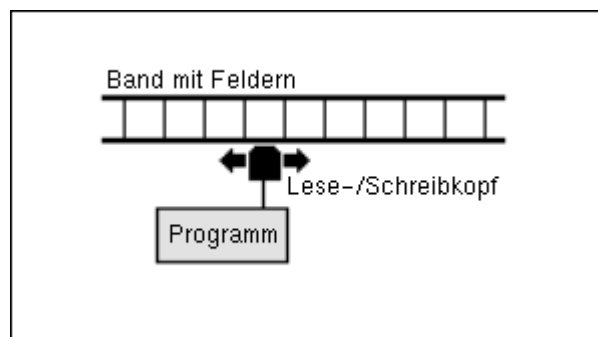
Im Sinne der computational theory of mind will man nicht nur die Semantik formalisieren, sondern mal will sie auch auf einen Computer implementieren. Man versucht sie also zu automatisieren.

Dazu sind zwei Bedingungen notwendig:

- 1) Keine Verwendung von semantischem Wissen, stattdessen soll die Anleitung in detaillierten Einzelschritten (Operationen) gehalten werden.
- 2) Die Anweisungen müssen in endlicher Zeit zu einem richtigen Ergebnis führen.

Diese zwei Bedingungen beschreiben auch einen Algorithmus. Darunter versteht man eine genau definierte Handlungsvorschrift zur Lösung eines Problems oder einer bestimmten Art von Problemen in endlich vielen Schritten.

Der Informatiker Alan Turing entwickelte in den 30er Jahren das theoretische Konzept der Turingmaschine. Seine These dazu lautet: Für jeden Algorithmus gibt es eine formal äquivalente Turingmaschine.



Was genau ist eine Turingmaschine? Diese Maschine besteht lediglich aus einem Lese bzw. Schreibkopf und einem langen Band mit Feldern. In jedem dieser Felder ist zu jedem Zeitpunkt genau ein Zeichen gespeichert. (0/1)

Zwischen den informationsverarbeitenden System (Lese-Schreibkopf) und dem Band besteht eine wechselseitige Beziehung. Die äußeren Zeichen auf dem Band werden vom System aufgenommen und verarbeitet, wonach eine Reaktion des Systems auf die Außenwelt zurückwirkt.

In jedem Schritt liest der Lese-Schreibkopf das aktuelle Zeichen, überschreibt dieses mit einem anderen Zeichen und bewegt sich dann ein Feld nach links oder rechts oder bleibt stehen. Welches Zeichen geschrieben wird und welche Bewegung ausgeführt wird, hängt von dem

an der aktuellen Position vorgefundenen Zeichen sowie dem Zustand ab, in dem sich die Turingmaschine gerade befindet. Dies wird durch eine zu der Turingmaschine gehörende Funktion definiert. Zu Beginn befindet sich die Turingmaschine in einem vorgegebenen Startzustand und geht bei jedem Schritt in einen neuen Zustand über.

Was kann man alles mit der Turing These berechnet werden? Turing war der Meinung, dass jede intuitiv berechenbare Funktion berechnet werden kann. D.h. Wenn ein Mensch zu einem bestimmten Ergebnis einer Rechenaufgabe kommt, kommt die Turing Maschine zum gleichen Ergebnis. Bei diesem Parallelismus von Mensch und Maschine handelt es sich um ein Modell der Berechenbarkeit. Wenn wir den Output der Maschine interpretieren, so erhalten wir eine wahre Aussage in der Mathematik.

Das Besondere an einer Turingmaschine ist, dass sie mit nur drei Operationen (Lesen, Schreiben und Kopf bewegen) die Probleme lösen kann, die auch von einem Computer gelöst werden könnten. Sämtliche mathematische Grundfunktionen wie Addition und Multiplikation lassen sich mit diesen drei Operationen simulieren. Darauf aufbauend kann man schließlich komplexere Programme simulieren.

Mensch vs. Computer

Die Frage die sich stellt ist, ob man das maschinelle Rechnen mit dem menschlichen Rechnen gleichsetzen kann. Die Turingmaschine liefert uns nur die Ergebnisse, welche dann aber von einem Menschen interpretiert werden müssen. D.h. Die Turingmaschine ist lediglich ein Modell der menschlichen Rechenfähigkeit. Um das maschinelle mit dem menschlichen Rechnen gleichsetzen zu können, müsste die Maschine selbstständig rechnen können und ein Verständnis für diese formalen Symbole entwickeln.

Zum Vergleich der Leistungsfähigkeit zwischen Mensch und Maschine gibt Alan Turing den so genannten Turingtest an, welcher ein Maß für die künstliche Intelligenz liefern soll. Intelligenz ist allerdings ein sehr dehnbarer Begriff, so beschreibt 1961 der Psychologie David Wechsler die Intelligenz eines Menschen als „zusammengesetzte oder globale Fähigkeit des Individuums, zweckvoll zu handeln, vernünftig zu denken und sich mit seiner Umgebung wirkungsvoll auseinander zu setzen“.

Der Turingtest besteht aus zwei unterschiedlichen Situationen. In der ersten Situation befinden sich eine Frau und ein Mann in einem abgegrenzten Raum. Die beiden haben über eine Tastatur und einen

Bildschirm Kontakt zu einer dritten Person, dem Fragesteller. Die Aufgabe des Fragestellers ist herauszufinden, wer von den beiden Personen der Mann und wer die Frau ist. Der Mann hat allerdings die Aufgabe das Verhalten einer Frau zu imitieren.

In der zweiten Situation wird nun der Mann durch eine Maschine ersetzt. Die Maschine hat den Turingtest bestanden, wenn sich der Fragesteller in der zweiten Situation von der Maschine genauso oft täuschen lässt, wie in der ersten Situation vom Mann.

Ein ähnliches Beispiel für diese Interaktion ist das von Joseph Weizenbaum entwickelte Computerprogramm namens Eliza, welches die Möglichkeit der Kommunikation zwischen Mensch und Computer zeigt. Eliza kann über verschiedene Skripten unterschiedliche Gesprächspartner simulieren. Bekannt geworden ist das Programm für die Simulation eines Psychiaters.

Das einfache Prinzip hinter dem Programm ist, dass Aussagen eines menschlichen Gesprächspartners in Fragen umformuliert werden, um so eine Reaktion auf die Aussagen zu erhalten.

Mensch: „Ich habe ein Problem mit meinem Nachbar.“

Eliza: „Warum haben sie ein Problem mit Ihrem Nachbar?“

Außerdem ist es in der Lage Schlüsselwörter zu erkennen.

Mensch: „Ich habe ein Problem mit meinem Vater.“

Eliza: „Erzählen Sie mir mehr von Ihrer Familie!“

So erkennt das Programm, dass Vater zum Oberbegriff Familie gehört und kann dadurch eine sinnvolle Antwort geben.

Das Kommunikationsverhalten der Versuchspersonen gegenüber dem Programm entsprach dem gegenüber einem menschlichen Gesprächspartner. Zum Teil waren die Versuchspersonen sogar davon überzeugt, dass der Gesprächspartner ein tatsächliches Verständnis für ihre Probleme zeigte. Bei etwas komplizierteren Aussagen oder unpassenden Fragen konnte der Gesprächspartner allerdings schnell als Programm enttarnt werden.

Die Problematik beim Turingtest ist dabei, auch wenn ein Computerprogramm den Turingtest besteht, ist das kein Beweis dafür, dass das Programm auch ein Verständnis für die verarbeiteten Symbole hat. Man sieht hier also wie wichtig der Aspekt der Semantik für intelligentes Verhalten ist.

John Searl

Der amerikanische Sprachphilosoph John Searle hat 1980 in einem Gedankenexperiment zu zeigen versucht, dass Computerprogramme natürliche Intelligenz nicht ersetzen können. Er nahm dabei ins Besondere Bezug auf das Verstehen von Symbolen. KI kann nämlich nur nach syntaktischen Prinzipien arbeiten und nie über inhärente Semantik verfügen. Syntaktisches Prinzip: In ELIZA, einem Programm zur simulierten Psychotherapie, wird dies unter anderem deutlich. Eliza reagiert auf Schlüsselwörter. Searl stellte die in behavioristischer Tradition stehenden Turing-Tests in Frage.

Searl nahm Gödels Theorem als Grundlage seiner Intention. Unvollständigkeitstheorem: Gilt für uns als denkende Wesen: Unser eigenes Denken kann nicht vollständig formalisiert werden, das heißt aber nicht automatisch, dass das Programm der KI nicht funktionieren kann. Schank Skripte: Zum Beispiel Ich gehe Einkaufen in den Supermarkt. Wir schließen daraus, dass ich Dinge mitnehme, an der Kasse bezahle und gehe. Schank bezieht sich auf Ausnahmesituationen, die eintreten können. Zum Beispiel, dass ich eine Abrechnungskarte der Lebenshilfe besitze, und automatisch eine Rechnung an meine Institution geschickt wird. Schank sagt nun, dass es unmöglich ist, ein künstliches System mit allen „Ausnahmesituationen“ zu füttern. Dieses Fehlen eines generellen Interpretationsschemas ist als so genanntes Frame- Problem bekannt.

Das Chinesische-Zimmer-Gedankenexperiment:

Kurz beschrieben geht es darum, dass eine Person, die kein Chinesisch versteht, teilnimmt. Nacheinander werden 3 Stapel mit chinesischen Schriftzeichen, die für den Probanden nur als Gekritzelt gedeutet werden können, in den Raum gereicht. Diese entsprechen (nach Schank) den Elementen „Skript“, „Geschichte“ und „Fragen“, ohne, dass dies die Person weiß. Zu den Paketen erhält der Proband Anweisungen, die Beziehungen zwischen den Stapeln herstellen, so zu sagen ein „Programm“. Aufgrund der Anweisungen werden Zettel wieder aus dem Raum gereicht, die Antworten auf die „Fragen“. Und obwohl richtige Dinge herauskommen, kann aus dem bloßen Hantieren mit bedeutungslosen Kritzelzeichen nicht daraus geschlossen werden, was sie bedeuten, oder dass der Proband chinesisch versteht. Im Gegensatz zu den Turing-Maschinen liegt es im Ermessen des Probanden, der wie eine Maschine funktioniert, und nicht des Beobachters, inwieweit ein Computerprogramm im buchstäblichen Sinne verstehen kann.

Daraus ergeben sich drei Axiome:

1. Computerprogramme funktionieren allein nach syntaktischen Prinzipien
2. Menschliches Verstehen erfolgt auf der Grundlage geistiger Inhalte (Semantik)

3. Semantik lässt sich nicht durch Syntax erklären

Daraus ergibt sich der Unterschied zwischen Computerprogrammen und menschlichem Verstehen: Intentionalität ist ein biologisches Phänomen und benötigt Symbole. Demnach vertritt Searl streng genommen ein Überbleibsel des Dualismus, in dem geistige Vorgänge als unabhängig von materiellen Prozessen betrachtet werden.

Die verkörperlichte und situierte Kognitionswissenschaft wiederum geht davon aus, dass Kognition nicht ohne Bezug auf einen spezifischen Körper (Verkörperlichung, Embodiment) und eine spezifische Umgebung (Situiertheit) erklärt werden kann. Diese Forderungen resultieren aus dem Zweifel daran, dass Kognition ein Prozess ist, der sich in einer Welt der abstrakten symbolischen Repräsentationen vollzieht, verhältnismäßig unabhängig von der genauen Sensorik, Motorik und dem zeitlichen Geschehen in der Außenwelt. Bekannte Vertreter dieser Auffassung sind Alva Noë, Susan Hurley, Evan Thompson, Francisco Varela und Kevin O'Regan. Im Rahmen der verkörperlichten und situierten Kognitionswissenschaft wird häufig eine Verknüpfung vom Gedankengut der Phänomenologie Maurice Merleau-Pontys und Edmund Husserls mit der klassischen analytischen Philosophie des Geistes angestrebt.

Diese verschiedenen vorgestellten Strömungen (Konnektionismus, Dynamizismus, Künstliches Leben, Situiertheit und Verkörperlichung) werden gerne unter dem Schlagwort *Neue KI (New AI)* zusammengefasst, da sie sich z. T. in ihren Forderungen und Annahmen überlappen. Allerdings können sie nicht als deckungsgleich betrachtet werden, da sie sich in vielerlei Weise in Prämissen, Konsequenzen und Anwendungen unterscheiden oder sogar widersprechen.

Roboter-Replik (Yale University):

Ein Computerprogramm wird in einen Roboter eingebaut, um über eine Camera und die motorische Steuerung von Greifarmen eine Einbettung in die Welt zu erreichen. Dies würde Searles Theorie widerlegen. Es sei, laut dem Team in Yale, genau diese Einbettung in die Welt, die Symbolketten eine Bedeutung verleihen. Searl unterstrich seinerseits wieder seine ursprüngliche Meinung, weil das was das Programm als Eingabe wiederum nur weitere bedeutungslose Symbole erhält.

Die einzige Lösung dieses Problems der KI aus heutiger Sicht wäre eine Möglichkeit zu finden, wie Systeme selbständig Symbole in ihrer Bedeutung und Deutung erlernen können. Womit wir beim so genannten Symbol Grounding Problem angekommen wären.

In Anlehnung an die Kritik von Searl an der starken KI stellt Stevan Harnad fest, dass so lange Zeichenketten ausschließlich anhand ihrer formalen Kriterien verarbeitet werden, nur eine abgeleitete Bedeutung eines externen Interpreten gegeben werden kann. Welche Möglichkeit sieht Harnad nun, dieses Problem zu umgehen? Anhand von Searles Gedankenexperiment schlägt er zwei Varianten der Herangehensweise vor: In der ersten Variante soll von einer Person, die bereits eine Sprache spricht, Chinesisch anhand eines Chinesisch-Chinesisch Wörterbuchs erlernt werden. Dieser Versuch wird zum Scheitern verurteilt sein. In der zweiten Variante wird mit der Situation eines Computerprogramms verglichen, das Chinesisch als *erste* Sprache erlernen soll. Nach Harnad gibt es nur eine Methode, um dieses Problem zu lösen. Die Symbole müssen irgendwie mit der Außenwelt verknüpft werden.

Biologische Grundlagen neuronaler Netze

Ein "**neuronales Netz (NN)**" entspricht, angelehnt an Anatomie und Physiologie des menschlichen Gehirns, einer großen Menge sehr einfacher Prozessoren, Neuronen, die vielfach über Synapsen verbunden sind. Ein Neuron beispielsweise steht in Verbindung mit über 10.000 weiteren Neuronen.

Ein Neuron besteht aus dem Zellkörper (Soma), dem Axon, den Dendriten und den Synapsen an den Endigungen der Dendriten und des Axons. Das Axon ist eine röhrenförmige Leitungsbahn, die die Signale zu anderen Nervenzellen überträgt, die Dendriten sind dünn verzweigte Ausläufer des Zellkörpers und dienen als Eingänge des Neurons. Sie stehen über die Synapsen mit den Axonen anderer Neurone in Verbindung. Die Form und Anordnung der Neuronen im ZNS (Länge sowie Zahl und Anordnung von Dendriten und Axon) ist sehr unterschiedlich, sie ist abhängig von ihrer jeweiligen Lage und ihrer Funktion im ZNS.

Übertragung auf datentechnische Verarbeitungsmodelle

Auch im Gehirn geht es natürlich physikalisch-chemisch zu. Warum sollte man das also nicht letztendlich künstlich nachbauen können. Das Nachbauen könnte aber so unendlich schwierig sein, dass es praktisch unmöglich ist. Das heißt, theoretisch, müsste man Systeme bauen können, die entweder geistiges Bewusstsein haben, oder sich zumindest so verhalten, dass sie ununterscheidbar sind von Wesen, die so etwas haben, nämlich uns. Das mag einfach sein, es mag sich aber auch als sehr schwierig oder völlig unmöglich im technischen Sinne erweisen.

Neuronale Netzwerke sind in Struktur und Funktionsweise ihren biologischen Vorbildern nachempfunden. Sie sind Systeme aus kleinen, untereinander verbundenen Verarbeitungseinheiten (= Prozessorelemente) und sind in der Lage, durch Anpassung der Stärke der Verbindungen zwischen den einzelnen Prozessorelementen, zu lernen. Strukturelle Ähnlichkeit und Lernfähigkeit sind aber nicht die einzigen Parallelen zu natürlichen neuronalen Netzwerken;

Das Gehirn ist ein sehr komplexes, nichtlineares, und parallel arbeitendes Informationsverarbeitungssystem. Es ist fähig, seine Komponenten so zu organisieren, dass es bestimmte Aufgaben, zum Beispiel Mustererkennung, sehr viel schneller ausführen kann, als die schnellsten bisher bekannten Computer.

Ein künstliches neuronales Netz versucht u.a. zu modellieren, wie ein biologisches Neuronennetz eine bestimmte Aufgabe oder Funktion erlernt; es durchläuft dafür einen Lernprozess. Diese Lernfähigkeit ermöglicht es einem neuronalen Netz, eine Aufgabe selbständig anhand von Trainingsbeispielen zu lernen, ohne dass dieses Netz dafür explizit programmiert werden muss.

»Sind neuronale Netze tatsächlich Modelle des menschlichen Gehirns?« (Anderson, 1989).

Diese Frage scheint durchaus berechtigt, angesichts der Tatsache, dass der Aufbau der Prozessorelemente und die Art der Verbindungen zwischen diesen Verarbeitungseinheiten, bei allen bisher entwickelten neuronalen Netzwerken - im Vergleich zu biologischen Systemen - eine mehr oder weniger starke Vereinfachung darstellt.

Obwohl die jüngst entwickelten Netzwerkmodelle den physiologischen Gegebenheiten mehr und mehr Rechnung tragen, scheint eine physiologisch exakte Nachbildung schier unmöglich zu sein. Man versucht, Strukturelemente eines Neurons in ein datentechnisches Verarbeitungsmodell zu überführen.

- Der Zellkörper fungiert als Informationsträger. Im einfachsten Fall werden 2 Zustände unterschieden. Erregt und nicht erregt.
- Das Axon als Weitervermittler des Erregungszustandes
- Die Synapse zur Bestimmung, wie sich die über ein Axon vermittelte Erregung auf andere Zellen auswirkt .

Das menschliche Gehirn ist jedoch weit komplizierter als diese einfachen Regeln. Es benötigt unter anderem noch Rezeptoren als Informationsträger und Effektoren zur Ausgabe.

Die vorgeschaltete Ansteuerung wird bei neuronalen Netzen als Eingabecodierung, die nachgeschaltete Ausgabe als Ausgabecodierung bezeichnet. **Eingabecodierung** heißt, die Fragen, den Sachverhalt, das Problem der Anwendung (entsprechend den Netztypkonventionen) zu verschlüsseln. Die Reaktion des Netzes ist dann wieder zu einer Antwort, einer Beurteilung oder einer Lösung zu dekodieren.

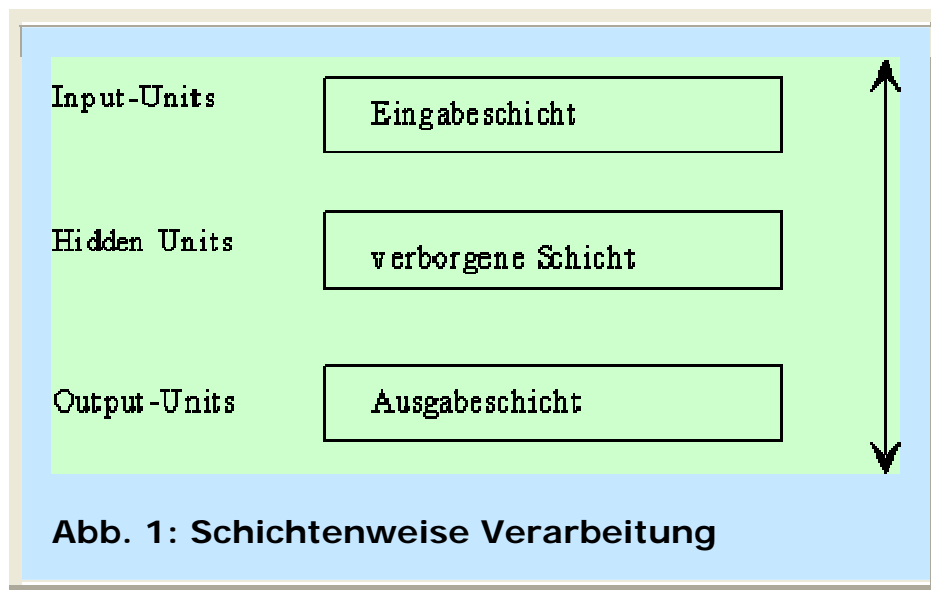
Allgemeiner Aufbau neuronaler Netze

Es gibt nicht das „neuronale Netz“ an sich. Bis jetzt gibt es einige unterschiedliche Modelle und Ansätze. Ein "künstliches neuronales Netz" besteht generell aus verschiedenen Elementen, Strukturen und Regeln.

Die grundlegenden Bestandteile sind:

- Verarbeitungselemente (Neuronen)
- eine Menge von Aktivitätszuständen
- eine Ausgabefunktion für jedes Element
- eine Verbindungshierarchie
- eine Regel zur Fortschaltung (Propagierung) der Aktivitätsmuster von einem Element zum nächsten
- eine Regel zur Aktivierung eines Elements durch anliegende Signale und Erzeugung eines neuen Aktivitätszustands
- eine Lernregel zur Modifizierung der Verbindungsgewichte
- eine Umgebung, in der das Netz arbeitet

Die Verarbeitung in solchen Netzwerken erfolgt in der Regel von der Eingabeschicht über verborgene Schichten zur Ausgabeschicht, gelegentlich aber auch von der Eingabeschicht direkt zur Ausgabeschicht oder in einem Feedback zwischen den einzelnen Schichten (hin und her).



Die einzelnen Schichten bestehen aus einer Anzahl von Prozessorelementen. Jede Schicht enthält mindestens ein Prozessorelement. Elemente einer Schicht sind meistens von der gleichen Art.

Ein Prozessorelement kann mit beliebig vielen anderen Prozessorelementen einer anderen Schicht ("inter-neuronlayer-connection") verbunden sein oder auch mit Prozessorelementen der gleichen Schicht (intra-neuronlayer-connection). Eine Verbindung von Prozessorelement_j nach Prozessorelement_i heißt Gewicht, neben variablen Gewichten gibt es auch feste Gewichte. Gewichte werden gewöhnlich durch reelle Zahlen (häufig im Intervall von -1 bis +1) dargestellt.

In vielen Fällen gibt es eine zusätzliche Schicht mit genau einem Element, das Bias genannt wird. Dieses Element hat den konstanten Wert 1 und hat nur Ausgänge, keine Eingänge. Mit Hilfe des Bias-Elements kann z.B. sichergestellt werden, dass bestimmte Prozessorelemente immer eine Eingabe ungleich Null erhalten.

Verbindungen zwischen Neuronen der gleichen Schicht sind häufig sowohl erregender als auch hemmender Art, während zwischen Schichten meistens nur erregende Verbindungen bestehen. Falls sich die Aktivierung von Schicht zu Schicht vorwärts ausbreitet, spricht man von Feedforward-Netzen. Es handelt sich um Feedback-Netze, falls die Aktivierung von nachfolgenden Schichten auch an vorgelagerte Schichten zurückgegeben wird

„Chaos“

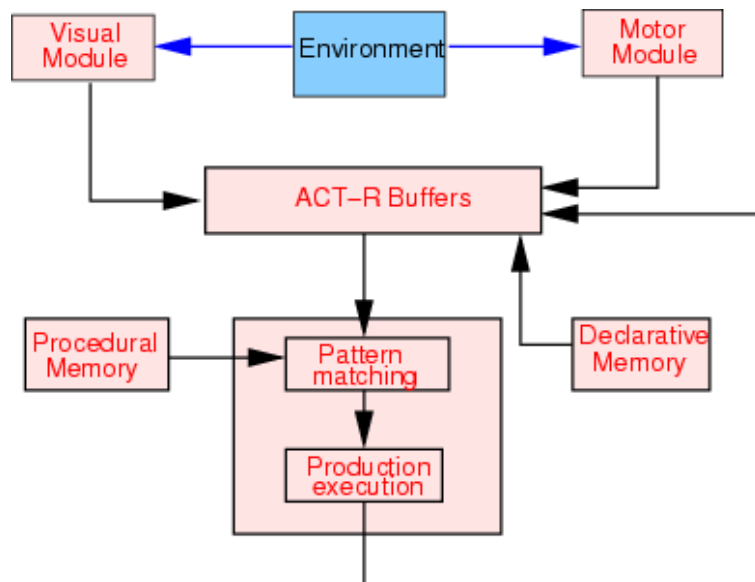
Von Chaos spricht man bei Phänomenen, die sich jeglicher Kontrolle und Berechenbarkeit entziehen, und die unbeherrschbar erscheinen. (umgangssprachlich)

„Chaostheorie“

Eine mathematische Theorie, die sich mit Systemen befasst, deren Verhalten langfristig unvorhersagbar und scheinbar regellos ist, obwohl ihre Komponenten durch eindeutige Gesetze beherrscht werden

Konnektionismus

Für neuronale Netze ist die Parallelität der Verarbeitung von Informationen charakteristisch. Dieser Ansatz wird als *Konnektionismus* bezeichnet, weil hier die Simulation kognitiver Leistungen auf dem Prinzip der Verknüpfung (Konnex) einer Vielzahl paralleler Einheiten beruht. Dieser Ausdruck taucht sowohl in der Forschung über Künstliche Intelligenz wie auch in der kognitiven Psychologie auf. erfolgreichsten kognitiven Architekturen sind ACT-R (Adaptive Control of Thought, ACT),



Quellen:

G. Dorffner (Hrsg.) (1990): „*Konnektionismus in Artificial Intelligence und Kognitionsforschung*“. Salzburg: Springer-Verlag.

Leidlmair, Karl: Artikel: Was ist Cognitive Science?

Kosko, B. (1992): *Neural networks and fuzzy systems: a dynamical systems approach to machine intelligence*. Englewood Cliffs, NJ: Prentice-Hall.

Schöneburg, E., Hansen N., Gawelczyk A.(1992): *Neuronale Netzwerke. Einführung, Überblick und Anwendungsmöglichkeiten*. Haar bei München: Markt& Technik Verlag.

Links:

Eliza: www.med-ai.com/models/eliza.html.de

ACTR: <http://act-r.psy.cmu.edu>