

# 1 Inhaltsangabe

1	Inhaltsangabe.....	1
2	Einleitung .....	2
3	Die Einstellung von Hubert Dreyfus und seine Kritikpunkte .....	3
3.1	Die Rolle des Körpers bei intelligentem Verhalten .....	4
4	Die phänomenologische Kritik an der KI .....	4
5	Ein Rückblick auf die Geschichte der KI.....	6
5.1	Der Paradigmawechsel nach Rodney Brooks .....	6
5.2	Schritte zur Komparativen Kognitiven Robotik.....	7
6	Die KI Forschung in der Sackgasse .....	13
6.1	Die KI als ein degenerierendes Forschungsprogramm .....	13
6.2	Das Relevanz- Problem.....	14
6.3	Das Frame- Problem.....	14
6.4	Das Problem des Alltagswissens.....	15
6.5	Die Mikrowelten .....	16
7	Abschließende Bemerkungen.....	17
8	Literaturverzeichnis.....	18

## 2 Einleitung

Der Ausgangspunkt für unser Referat war ein bis jetzt noch unveröffentlichter Artikel von Hubert L. Dreyfus mit dem Titel „Why Heideggerian AI failed and how fixing it would require making it more Heideggerian“. Zu allererst möchten wir die Person Hubert Dreyfus etwas näher vorstellen.

Hubert L. Dreyfus wurde 1929 geboren und ist ein US-amerikanischer Philosoph. Er lehrt an der Universität Kalifornien und bis 1968 am Massachusetts Institute of Technology. Dreyfus' Arbeitsschwerpunkte sind Phänomenologie, Existenzialismus, Philosophie der Psychologie und Literatur und Philosophische Tragweite der Künstlichen Intelligenz (KI).

Dreyfus ist besonders bekannt durch seine Kritik an der Künstlichen Intelligenz. Im Jahr 1965 publizierte er „Alchemy and Artificial Intelligence“, in dem er der Grundannahme, dass Intelligenz aus Manipulation von physikalischen Symbolen entsprechend formaler Regeln besteht, widerspricht und argumentiert, dass das KI-Forschungsprogramm zum Scheitern verurteilt sei. Dreyfus arbeitete 1965 bei der Rand Corporation und war er Gastprofessor an der Universität Frankfurt und am Hamilton College.

Ein weiterer Begriff, den wir hier noch einführen wollen, ist das MIT, da auch es eine wichtige Rolle im Artikel spielt.

Das Massachusetts Institute of Technology (MIT), gegründet 1861 ist eine weltbekannte Universität in Cambridge (Massachusetts) in den USA. Das MIT gilt als eine weltweit führende Universität im Bereich von technologischer Forschung und Lehre. Es ist eine private, nicht-konfessionelle technische Universität, die als erste Chemie-Ingenieure ausbildete und die Wirtschafts-, Sozial- und Geisteswissenschaften in die Ingenieurausbildung einbezog. Die Hochschule liegt am Charles River in Cambridge, direkt gegenüber von Boston und stromabwärts von der Harvard-Universität. Neben dem Schwerpunkt Naturwissenschaften und Technik hat das MIT auch Institute für Philosophie, Betriebswirtschaft, Linguistik und Anthropologie.

Unter den bekanntesten Forschungsinstituten sind zu nennen:

Das „Lincoln Lab“, das „Research Lab of Electronics“, das „Lab for Information and Decision Systems“, „Computer Science and Artificial Intelligence Lab“ (!), das „Media Lab“, und die „MIT Sloan School of Management“.

(vgl. <http://www.wikipedia.de>)

### **3 Die Einstellung von Hubert Dreyfus und seine Kritikpunkte**

Einige der wichtigsten Werke von Dreyfus sind:

- Alchemy and Artificial Intelligence (1965)
- What Computers Can't Do: The Limits of Artificial Intelligence (1972)
- What Computers Still Can't Do: A Critique of Artificial Reason (1979)
- Mind Over Machine (1986)
- On the Internet (2001)

Sein Buch über die KI-Forschung „What computers can't do“ (1972) wurde seinerzeit laut kritisiert und doch still gelesen.

Dreyfus kritisierte schon sehr früh die Annahmen und Methoden der KI-Forscher und wurde auch vor allem dadurch bekannt. Seine Überzeugung basierte u.a. auf der philosophischen Annahme, dass die menschliche Intelligenz nicht durch eine „einfache“ Rechenmaschine wie dem Computer simuliert werden kann. Somit widerspricht er der Grundannahme, dass Intelligenz aus Manipulation physikalischer Symbole entsprechend formaler Regeln besteht. Dreyfus sah schon lange Probleme in der künstlichen Abbildung des menschlichen Alltagswissens. Er prognostizierte früh eine Entwicklung der KI in eine Sackgasse und sah eine Lösung nur in einer grundsätzlich anderen Modellierung, vor allem in holistischen Systemen. Auf der Ebene der Methodik kritisierte er die KI-Forscher aufgrund ihrer mangelnden Selbsteinschätzung und übertriebenen Versprechen.

Neben dem Aufzeigen der vermeintlichen Fehler zeigt er in seinem Buch auch alternative Annahmen auf, die in der KI-Forschung betrachtet werden sollten, wie z.B. die Rolle des Körpers, die veränderliche Relevanz, und das „knowing- that“ vs. „knowing- how“.

Im Folgenden wollen wir noch kurz näher darauf eingehen, was Dreyfus über die Rolle des Körpers sagt.

### **3.1 Die Rolle des Körpers bei intelligentem Verhalten**

Hubert Dreyfus stellte bereits 1972 das cartesianische Computer- Paradigma in Frage und kritisierte auch die Anwendung desselben. Er berief sich stattdessen auf die französische Phänomenologie rund um Maurice Merleau- Ponty. Seine Kritik am Cartesianischen Konzept beinhaltete z.B., dass dieses Modell von einer passiven Informationsaufnahme ausgeht und den Körper als eher hinderlich für die Intelligenz betrachtet. Laut seinem phänomenologischen Konzept gibt es eine sensorische Informationsaufnahme, und der Körper wird als ein wesentlicher Bestandteil von Intelligenz betrachtet. Dreyfus geht sogar so weit zu behaupten, dass nicht- formalisierbare Formen der Datenverarbeitung wie z.B. die Mustererkennung nur von Wesen bewältigt werden können, die mit einem Körper ausgestattet sind. Weiters betont Dreyfus einen wesentlichen Unterschied zwischen Mensch und Maschine bzgl. der Verwertung der Inputs: Menschen verfügen über einen inneren Horizont, sie erkennen Stimuli, indem sie diese mit ihren Erfahrungen in Beziehung setzten. Dagegen kann eine Maschine die Inputs lediglich mit einer vorprogrammierten Liste auf Entsprechungen hin überprüfen.

(vgl. <http://www.ifi.unizh.ch/groups/ailab/teaching/semi2000/Folien.pdf>)

## **4 Die phänomenologische Kritik an der KI**

Die Heideggersche Phänomenologie und die KI wurden durch Dreyfus' Veröffentlichung *Alchemie und KI* bereits in den 60er Jahren miteinander in Verbindung gebracht. Doch erst in den 80er Jahren fand diese Verbindung eine breitere Fachöffentlichkeit. Das größere Interesse in der Fachwelt ist unter anderem durch Winograds Abkehr vom traditionellen Forschungsprogramm der KI und seinem Buch *Understanding Computers and Cognition* zu erklären. Im Wesentlichen wird von Dreyfus, Winograd & Flores und anderen die folgende Kritik an dem klassischen, dem symbolischen Ansatz der künstlichen Intelligenz angebracht:

Es wird argumentiert, dass für symbolmanipulierende Systeme Informationen bloß *vorhanden* sind. Maschinen müssen die Vorhandenheit einer fixierten 'Welt' repräsentieren und können nur die repräsentierenden Symbole nach starren, festen Regeln manipulieren. Im Gegensatz dazu macht aber im menschlichen Denken und Handeln die „Zuhandenheit“ den weitaus größten und wichtigsten Teil aus. Eine externe Welt wird gar nicht erst bewusst gemacht. Menschliches Wissen ist zu einem großen Teil nur implizit im menschlichen Handeln, in dessen spezifischer Struktur enthalten - und nicht symbolisch repräsentiert oder repräsentierbar, so Dreyfus.

Diese Beobachtung reduziert Dreyfus zunächst auf die Unterscheidung zwischen *knowing-how* und *knowing-that*. Mit *knowing-how* ist das Handlungswissen gemeint, das sich nur in der tatsächlich durchgeführten Handlung *zeigt*. Dieses Wissen ist - da es sich nicht durch Symbole repräsentieren lässt - einem Formalismus auch nicht mitteilbar - so Dreyfus. Demgegenüber steht das Wissen von der Art des *knowing-that*, welches durch Symbole repräsentierbares Wissen meint. Dreyfus behauptet also zunächst die Wesensverschiedenheit von *knowing-how* und *knowing-that* und die entsprechende Dichotomie die sich damit über den verschiedenen Arten von Wissen aufspannt.

Dies mag überraschen, da Dreyfus Heideggers *Sein und Zeit* im Wesentlichen mit Wittgensteins Spätphilosophie gleichsetzt. Dabei weist er darauf hin, dass der Ansatz der traditionellen KI eigentlich Wittgensteins Frühphilosophie aus dem *Tractatus* entspricht. Wittgenstein macht jedoch bereits im *Tractatus* auf die Unterscheidung zwischen sprachlich Beschreibbarem und nicht Beschreibbarem aufmerksam:

Es gibt allerdings Unaussprechliches. Dies *zeigt* sich, es ist das Mystische.

Weiterhin behauptet Dreyfus für die nicht symbolisch repräsentierbare Wissensart des *knowing-how*, dass sie eines nicht durch algorithmische Regeln beschreibbaren Inhaltes ist. Das Verhalten ist nur im philosophischen Sinne regelgeleitet - folgt also nur Regeln mit nicht näher bestimmbar Ausnahmen. Daraus schließt Dreyfus letztendlich, dass menschliches Intelligenzverhalten nicht algorithmisierbar ist - da es auch wesentlich Wissen von der nicht algorithmisch beschreibbaren Art des *knowing-how* involviert. (vgl. [www.cse.unsw.edu.au/~achim/Research/Philosophie/node62.html](http://www.cse.unsw.edu.au/~achim/Research/Philosophie/node62.html) - 8k)

## 5 Ein Rückblick auf die Geschichte der KI

Die KI Forschung hat ursprünglich einen verheißungsvollen Beginn mit den Arbeiten von Newell & Simon in den frühen 60 Jahren und es kommt zum Aufstieg zu einem florierendem Forschungsprogramm durch die Erfolge der Mikrowelten (ca. 1970). Dadurch lässt sich Marvin Minsky zu der Vorhersage hinreisen, dass wir schon innerhalb der nächsten Generation intelligente Computer haben werden. Doch plötzlich kommt es zu Schwierigkeiten, z.B. wie kann ich das Alltagswissen oder die Mustererkennung in solche Programmierungen integrieren.

Laut Dreyfus geht Entwicklung hin zu einem degenerierenden Forschungsprogramm.

### 5.1 Der Paradigmawechsel nach Rodney Brooks

Rodney Allen Brooks (\* 30. Dezember 1954 in Adelaide) ist Direktor des *Computer Science and Artificial Intelligence Laboratory* (CSAIL) am MIT.

Nach dem Studium der Mathematik an der Flinders University of South Australia hat Rodney Brooks 1981 seinen Doktor in Informatik an der Stanford Universität abgelegt. Seine Forschungsschwerpunkte liegen einerseits der Konstruktion einfacher, aber intelligenter Roboter, die sich in dynamischen und komplexen Umgebungen frei bewegen können, sowie andererseits in der Erforschung komplexer menschlicher Intelligenz durch die Konstruktion humanoider Roboter. Er ist damit ein klassischer Vertreter der verkörperten KI.

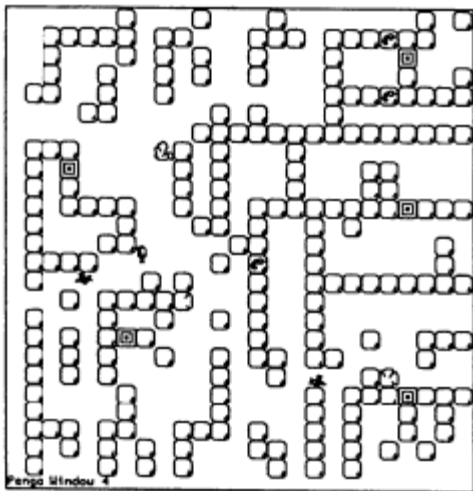
Rodney Brooks Ansatz zur Konstruktion von Robotern, die sich in dynamischen Umgebungen flexibel bewegen können liegt in der Unterteilung des Verhaltensrepertoires in viele parallel berechnete Verhaltensmuster, von denen dann das jeweils adäquate automatisch ausgewählt wird. Anstatt in serieller Art und Weise verschiedene Berechnungsböcke in Reihe zu schalten (Wahrnehmung, Kognition, Handlung) konstruiert der Roboter keine Repräsentation seiner Umgebung, sondern handelt direkt auf seinen Sensordaten. Für jedes Verhaltensmuster ist jeweils nur ein Bruchteil der eingehenden Sensordaten relevant, wodurch die Komplexität der Berechnung stark eingeschränkt ist. Die Auswahl des tatsächlich ausgeführten Verhaltens ist durch die Sensordaten bestimmt, wobei nicht zwischen Selbstwahrnehmung (z.B. Batteriestatus) und Wahrnehmung der Welt unterschieden wird. Erreicht ein Verhaltensmuster seine aktivierungsschwelle unterdrückt es die Ausführung aller

anderen Verhaltensmuster. Eine solche Architektur erlaubt Reaktionen in Echtzeit, ist allerdings nicht lernfähig (vgl. [www.wikipedia.at](http://www.wikipedia.at)).

Mit Brooks kommt es zu einem Paradigmawechsel vom GOFAI (Good Old Fasioned AI) zu dem Ansatz des Embodied Cognitivistic. Der alte Ansatz, der das Input oder die wahrgenommen Umwelt als abstrakte Symbolwelt sieht, wird abgelöst durch die Ansicht, dass der Input oder die wahrgenommen Umwelt durch sensorische Stimulation erfolgt. Auch wird in dieser neuen Sichtweise die Informationsverarbeitung als Interaktion mit der Umwelt gesehen und nicht als einfaches Input/Output – Modell.

Brooks Ansatz ist nach Dreyfus ein wichtiger Fortschritt, aber Dreyfus kritisiert, dass Brooks Roboter nur auf bestimmte Charakteristika der Umwelt reagieren, nicht auf den jeweiligen Kontext oder auf sich verändernde Bedeutungen. Brooks nennt sie passender Weise „Animats“. Und noch ein Punkt, der kritisiert wird ist, dass Brooks sich über das „Lernen“ keine Gedanken macht.

## 5.2 Schritte zur Komparativen Kognitiven Robotik



### **Pengo : Ein Modell *menschlicher* Leistungen im Rahmen der Kognitiven Robotik**

Pengo – ein Videospiel, das von einem Computerprogramm gespielt werden kann (aus (Agre 1988 : 196))

**Philip Agre** und **David Chapman** schrieben Programme, die Videospiele spielen können. Wir wollen ihre Ideen nochmals im Kontext der Überlegungen, die wir inzwischen angestellt haben, betrachten (die Darstellung im folgenden orientiert sich an (Agre 1988)) : die Aufgabe des Programms („Pengi“), das wie gesagt die Rolle und Perspektive des Videospiele-Spielers (nicht der Spielfigur!) übernimmt, ist es, das Videospiele („Pengo“) erfolgreich zu spielen. Dazu muss der Pinguin durch das Labyrinth von Eisblöcken gesteuert werden und spezielle Blöcke nebeneinanderreihen. Gelingt ihm das, so hat er das Spiel gewonnen. Man kann die Eisblöcke schießen; dann fliegen sie soweit entlang der Schussrichtung, bis sie auf einen anderen Eisblock oder den Spielfeldrand treffen. Die Bienen, die auf dem Spielfeld umherfliegen, sind gefährlich : wenn sie den Pinguin berühren oder wenn sie einen Eisblock auf ihn schießen, verliert er sein Leben. Er kann sich wehren, indem er seinerseits Eisblöcke auf die Bienen schießt.

Warum programmierten Agre und Chapman Programme, die solche Videospiele spielen können? Nun: Anders als etwa beim Lösen von Denksportaufgaben oder beim Schachspielen (das sind und waren typische Untersuchungsgegenstände der Kognitionswissenschaft) kommt es hier darauf an, schnell und korrekt auf sich rasch verändernde Situationen zu reagieren. Da bleibt meist keine Zeit für *Sense – Model – Plan – Act* ! Das ist typisch auch für Alltagssituationen des Menschen (vgl. z.B. Autofahren oder Sich- in- einer- Menschenmenge- Bewegen). Durch Introspektion und Videoaufnahmen in Spielzentren kamen Agre und Chapman zu der Hypothese, dass (erfahrene!)Videospiele über ein Repertoire von Verhaltensweisen verfügen, die von spezifischen Situationsreizen ausgelöst werden, ähnlich wie die Roboter der Verhaltensbasierten KI.

Zu diesen Verhaltensweisen zählen auch typische Blickbewegungen über das Spielfeld. Sie versuchten, ein realistisches visuelles System zu implementieren, das auf der kognitionswissenschaftlichen Theorie der *Visual Routines* nach Shimon Ullman aufbaut (Ullman 1984). Das Programm blickt wie ein Roboter auf das Spielfeld auf dem Bildschirm und verwendet *nicht* die abstrakte Repräsentation des Spielfelds an anderer Stelle im Rechner. Relevante Elemente des Spielfelds (etwa Feinde) werden (in Anlehnung an Aufmerksamkeitseffekte)durch „visuelle Marker“ herausgehoben und verfolgt (→*visual tracking*). Die Ergebnisse der *Visual Routines* lösen Verhaltensweisen aus (z.B. Flucht) und starten ggf. neue solche Routinen (z.B. Suche nach einem sicheren Fluchtweg in der sich dynamisch verändernden Szenerie). Pengis *visuelles System* basiert wie gesagt auf Ullmans



„*Visual Routines*“, die z.B. dazu dienen festzustellen, ob zwei Objekte „auf einer Linie“ sind, ob die Spielfigur in einem nach allen Seiten abgeschlossenen Gebiet steht, was sich in einer Bewegungsrichtung befindet usw., und um „Marker“ auf der gesehenen Szene zu platzieren, um damit ggf. Objekte zu verfolgen. Pengis *motorisches System* besteht nur aus einfachen Richtungsanweisungen und einer Operation zum Schießen der Blöcke, entsprechend einer Joystick-Steuerung. Dazwischen liegt das *zentrale System*, das die Form eines simulierten elektronischen Schaltkreises hat. Das zentrale System wertet permanent die Ergebnisse der *Visual Routines* aus (z.B. „Hat sich das Objekt mit Marker 2 bewegt? j/n) und initiiert neue Routinen (z.B. „Finde die Biene, die dem Pinguin am nächsten ist.“). Die Ergebnisse, die die *Visual Routines* liefern, sprechen verschiedene Teile des zentralen Systems an. Das sind die Verhaltensweisen, auf die die aktuelle Situation „passt“. Sie berechnen Vorschläge ("*Arguments*") für die weitere Steuerung des Pinguins; diese Vorschläge aktivieren ggf. weitere Schaltkreise, die Gegenvorschläge machen usw. (das nennt Agre "*Running Arguments*"). Die in dieser „Debatte“ zuletzt gemachten Vorschläge „gewinnen“ und werden in der Motorik umgesetzt. Sie standen in der durch die Programmierer festgelegten (!) Hierarchie weit oben.

Betrachten wir nun Pengi in Aktion!

Wieso gerade diese Form gewählt wurde, ist unklar. Vielleicht liegt das daran, dass die Autoren als Mitarbeiter des MIT bestens mit elektronischen Schaltkreisen und ihrer Simulation vertraut waren. Hier zeigt sich wieder, dass die Programmierstrategien der Implementierungsebene arbiträr sind.

### **Beispiel: Pengi killt eine Biene**

Pengi muss den Pinguin erkennen (er wird mit Marker 0 markiert), die Biene (sofern es keine „relevanteren“ Bienen gibt), den Eisblock, und dass Biene und Eisblock „auf einer Linie“ sind. Marker 1 markiert immer die „im Moment relevanteste“ Biene und Marker 2 sucht und platziert sich auf weitere sich bewegende Objekte, um zu sehen, ob sie ggf. noch „relevanter“ (vor allem: „noch näher“) sind. Die verschiedenen Marker sind hier durch geometrische Formen angedeutet (vgl. Agre 1988: §C4e).

Pengi sucht ständig nach Eisblöcken, mit denen es ggf. Bienen killen könnte (ohne selbst in die Schusslinie zu geraten!); ein geeigneter Eisblock wird mit Marker 3 markiert; die Schusslinie wird mit einer visuellen Routine ermittelt und der Pinguin entsprechend bewegt: erst so, dass er auf einer Linie mit dem Eisblock ist; ist er auf einer Linie, wird eine

Bewegung auf den Eisblock zu initiiert (soweit die aktuell wahrgenommene Situation nicht eine andere Aktion dringender macht); dabei werden die Konditionen „die- Biene- ist- auf- einer- Linie- mit dem- Eisblock“, „es-gibt-keine-andere-Biene-die-mir-im-Augenblick- gefährlich-werden-könnte“etc. ständig parallel abgeprüft und aktivieren ggf. geeignete Aktionen. Wenn die Bedingungen keine andere, dringendere Aktion initiieren, schießt Pengi den Eisblock auf die Biene.

### **Pengi als Beispiel eines Systems der Verhaltensbasierten KI**

Anhand des Pengi- Systems können wir uns noch einmal die eingangs ausgeführten Charakteristika der Verhaltensbasierten KI vor Augen führen. Es stellen sich einige zentrale Fragen:

#### **Wieso kommt Pengi ohne ein Weltmodell aus?**

Weil Pengi „Deiktische Repräsentation“ verwendet, d.h.:

- Das System initiiert genau und nur die visuellen Operatoren, die die für die Steuerung des Pinguins relevanten Merkmale aus der Spielszene extrahieren. Es findet daher keine Konstruktion eines „Weltmodells“ statt, in dem z.B. alle Eiswürfel unterschieden würden.
- Die Objekte und ihre Relationen in einer Spielszene werden auf diese Weise relativ zu ihrer Bedeutung für den Spielverlauf individuiert. (analog zu „die- Biene- die- mir- am- nächsten- ist“, „die- Biene- auf- die- ich- einen-Eisblockschießen- kann“ und auch „die- Biene- die- mir- am- nächsten- ist- kann- einen-Eisblock- in meine- Richtung- schießen“, ...)
- Deshalb werden Aktionen „automatisch“ auf neue Elemente einer Szene übertragen, wenn sich die Spielszene dynamisch ändert (z.B. wird ggf. eine andere Biene zu „die- Biene- die- mir- am- nächsten- ist“); das ist *verhaltensbasierte Abstraktion, ohne Variablen!* (vgl. die berühmte Kritik von Fodor und Pylyshyn an konnektionistischen Modellen, weil diese keine Variablen haben (Fodor / Pylyshyn 1988), vgl. dazu auch (Fetzer 1992) und (Goschke / Koppelberg 1990, 1991))
- Wichtige Elemente der Spielszene werden durch Marker gekennzeichnet und ggf. verfolgt. (Es stellt sich die Frage, inwiefern sich die von Agre und Chapman vermittels *Visual Routines* realisierten visuellen Suchprozesse, die Analoga von Gedächtnis- und Aufmerksamkeitseffekten sein könnten, auch bei Menschen empirisch nachweisen lassen.

- Die Vorhersage geschieht durch „**Visualisierung**“ statt durch „Simulation in einem Weltmodell“: wenn Pengi wissen will, in welcher Richtung ein Eiswürfel gleiten wird, wenn er jetzt angestoßen würde, verfolgt es die entsprechende Richtung mit einer visuellen Routine (und entdeckt dort z.B. eine Biene).

### **Wieso kommt Pengi ohne Planung aus?**

Weil Pengi „*Routinen*“ realisiert, die in einem „*Running Argument*“ in jedem Augenblick neu auf die wahrgenommene Szene reagieren und in Aktionen umgesetzt werden, d.h.:

- Pengi „weiß“ schon alle Spieltricks (in Form von Verhaltensweisen / „Routinen“), auf die Agre und Chapman gekommen sind. Diese „Routinen“ sind Aktionen, die durch wahrgenommene Elemente der Szene angestoßen werden. Diese Aktionen (und die Veränderung der Szenerie, die unabhängig von Pengis Handlungen geschieht) realisieren dann wiederum eine neue Situation, in der neue Handlungen angestoßen werden (z.B. den Pinguin hinter einen Eisblock bewegen; wenn er hinter dem Eisblock steht, kann er ihn auf eine Biene schießen, sofern die Situation inzwischen nicht eine andere Handlung dringender macht).
- Pengis Routinen sind in einer vom Konstrukteur festgelegten Hierarchie so verdrahtet, so dass idealerweise zu jedem Zeitpunkt die jeweils wichtigste Aktion in die Tat umgesetzt wird.
- Pengi kann in begrenztem Umfang „in die Zukunft schauen“, indem es den möglichen weiteren Spielverlauf „visualisiert“ (durch visuelle Routinen, s.o.).
- Pengi hat sein Gedächtnis in der Welt, nicht im Kopf (!) (bis auf ein paar Ausnahmen durch verzögerte Signale in der internen Maschinerie), es benutzt die Welt als *externes Gedächtnis*, d.h. : wenn es „eine Handlung“ wegen einer Änderung in der Spielszene „unterbrechen“ muss (z.B. um einem heransausenden Eisblock auszuweichen), dann kann es hinterher wieder „zu dieser bzw. einer ähnlichen Handlung zurückfinden“, wenn die Spielszene (inklusive der Marker!) wieder eine „ähnliche Konstellation“ aufweist und dann wieder „die gleiche bzw. eine ähnliche Handlung“ anstößt. (Beispielsweise könnte Pengi mit seiner Spielfigur auf einen Eisblock zulaufen, um ihn auf eine Biene zu schießen, dann aber einer anderen Biene ausweichen. Ist die erste Biene noch immer auf einer Linie mit dem Eisblock, wird wieder

dieselbe Verhaltensweise ausgelöst und Pengi wird sich nach dem Ausweichen wieder daran machen, auf diese erste Biene zu schießen.)

Pengis komplexe Handlungsabläufe entstehen also so wie nachgewiesenermaßen oft bei Tieren (vgl. z.B. (Hinde 1987)), aber auch beim Menschen: auch dort führt bei vielen Handlungen ein Handlungsmuster zu einer neuen Konstellation in der Welt, die das folgende Handlungsmuster anstößt usw.

### **Der Bezug zur Kognitionswissenschaft**

Agre und Chapman postulieren, dass ähnliche Prozesse wie in Pengi auch in uns ablaufen, wenn wir die erstaunliche Performanz unserer „Alltagsintelligenz“ zeigen. Alltägliche Handlungen dienen nicht dem Lösen von immer neuen Problemen in einer unbekannt Welt, sondern sie bestehen überwiegend aus der routinierten Interaktion mit einer vertrauten Umwelt, die meist (durch uns selbst, durch soziale Konventionen, aber vielleicht auch nach einem allgemeinen Naturprinzip ?!) so strukturiert ist, dass wir in ihr ohne „Weltmodelle“, „Pläne“ und „Suchen im Lösungsraum“ auskommen. Die Welt ist meist so strukturiert, dass ähnliche Handlungen in einer ähnlichen Situation wieder zu einem ähnlichen Erfolg führen und dass ein Fehler (oder auch das Ausprobieren neuer Handlungen) nicht zum Tode führt und die Konsequenzen wieder rückgängig gemacht werden können. Begriffe wie „Weltmodelle“, „Pläne“ (bzw. „Algorithmen“ und „Heuristiken“), „Suchen im Lösungsraum“ sind bloß der untaugliche Versuch, die Abläufe in klassischen KI Systemen auf die Kognition des Menschen zu projizieren. Das ist eine entscheidende Einsicht aus den Programmen von Agre und Chapman !

(vgl. <http://www.ikw.uniosnabrueck.de/~roul/3%20%20Schritte%20zur%20Komparativen%20Kognitiven%20Robotik.pdf>)

## 6 Die KI Forschung in der Sackgasse

### 6.1 Die KI als ein degenerierendes Forschungsprogramm

Hubert Dreyfus hat die Künstliche Intelligenz häufig als ein degenerierendes Forschungsprogramm bezeichnet. Das liegt zum einen schon allein daran, dass sie sich laut ihm in einigen wesentlichen Punkten von einer richtigen Wissenschaft unterscheidet.

Zum Beispiel erwähnt Dreyfus immer wieder, dass die KI unfähig ist, sich in adäquater Weise mit ihren Fehlern zu beschäftigen und in weiterer Folge daraus zu lernen. Fehlschläge hat es in der Geschichte der KI zur Genüge gegeben, doch in den seltensten Fällen wurden Protokolle darüber veröffentlicht. In einer Wissenschaft gehört aber genau das zum Erkenntnisprozess dazu, da man teilweise gerade aus Fehlern wieder wichtige Dinge lernen kann.

Um beim Thema Fehlschläge zu bleiben: Dreyfus betont weiters, dass die KI- Forscher stets lahme Entschuldigungen für Fehlschläge verwendet haben. Auch hierdurch wird wieder deutlich, dass einer Konfrontation mit den eigenen Fehlern stets ausgewichen wurde.

Weiters erwähnt Dreyfus, dass die gesamte KI von einem übertriebenen Optimismus gekennzeichnet ist. Scheinbar fehlt hier ganz stark der Bezug zur Realität, stattdessen herrschen großartige Versprechungen und waghalsige Prognosen vor, wie dies auch bei Marvin Minsky sehr häufig der Fall war.

Als letzten Punkt nennt Dreyfus, dass es den KI- Forschern an theoretischen Vorüberlegungen mangelt. Eine Problemstellung wird nicht im Vorhinein theoretisch durchdacht, sondern einfach mittels der neuesten technischen Verfahren ausprobiert. Für den sonst in der Wissenschaft üblichen Prozess der Theorienbildung scheint in der KI wenig Zeit und Raum zu sein.

Dreyfus vergleicht die KI- Forschung des Öfteren mit der Alchemie, einem alten Zweig der Naturphilosophie, der im 17./ 18. Jahrhundert nach und nach von der Chemie und der Pharmakologie abgelöst wurde, und ebenfalls als degenerierendes Forschungsprogramm angesehen werden kann.

Ein weiterer Grund, warum sich die KI in der Sackgasse befindet, liegt darin, dass es ihr bis jetzt noch nicht möglich war und wahrscheinlich auch niemals möglich sein wird, bestimmte Probleme zu lösen, die für uns Menschen überhaupt nicht schwierig sind. Dazu zählen u.a. das Relevanz- Problem, das Frame- Problem, und das Problem des Alltagswissens.

(vgl. Dreyfus, Artikel und Dreyfus, 1985).

## **6.2 Das Relevanz- Problem**

Im Relevanz- Problem liegen zwei Schwierigkeiten verborgen: Zum einen muss man die Fakten, die möglicherweise relevant sind, einschränken (auf Grund der unendlich großen Anzahl an Tatsachen) und gleichzeitig aber die Allgemeingültigkeit erhalten. Zum anderen soll man die in einer bestimmten Situation tatsächlich relevanten Fakten auswählen können.

Menschen haben mit dieser Problematik überhaupt keine Schwierigkeiten, hingegen kann ein Computer das Relevanz- Problem überhaupt nicht lösen. Dreyfus erklärt diese Tatsache mit der Philosophie von Heidegger, der den Begriff „in- der- Welt- sein“ geprägt hat. Damit ist gemeint, dass wir Menschen in der Welt zuhause sind, es ist schon immer unsere Welt und wir haben sie bequem um uns herum angerichtet. Alles ist da wo wir es brauchen und hat dadurch für uns eine spezielle Bedeutung.

(vgl. Dreyfus, 1985)

## **6.3 Das Frame- Problem**

Das Frame- Problem ist ein Spezialfall des Relevanzproblems. Die zentrale Annahme dabei ist, dass ein Computer den derzeitigen Zustand der Welt simuliert. Wenn sich nun etwas in der Welt ändert, steht man vor dem Problem, dass der Computer nicht weiß, welche Fakten sich geändert haben und welche gleich geblieben sind.

Minskys Lösungsansatz zum Frame- Problem besagt, man sollte einfach die wichtigsten Situationen mit Auflistung der normalerweise relevanten Fakten programmieren. Eine solche Auflistung wird auch in der allgemeinen Psychologie als „Frame“ bezeichnet. Aber um den entsprechenden Frame für eine bestimmte Situation auswählen zu können, müsste der Computer auf einen hierarchisch übergeordneten Frame zugreifen, der die Merkmale enthält, wann der betreffende Frame zu verwenden ist. Aber auch zur Auswahl des übergeordneten

Frames würde wiederum ein noch weiter übergeordneter Frame benötigt, usw. Man stünde also vor einer endlosen Kette von Rückgriffen auf Frames, denn es gibt keinen, der an oberster Spitze dieser Hierarchie stehen könnte, da es keine endgültig richtigen und immer gültigen Tatsachen gibt. Somit kann ein Computer auch nicht auf diese Weise programmiert werden, und wird deshalb das Frame- Problem nicht lösen können. Auch in diesem Fall verhält es sich wie beim Relevanz- Problem: wir Menschen haben damit überhaupt keine Schwierigkeiten.

(Vgl. Dreyfus, Artikel)

## 6.4 Das Problem des Alltagswissens

In der Philosophie gibt es dafür leider keinen einheitlichen Terminus, selbst Dreyfus benutzt unterschiedliche Begriffe, meistens jedoch den Begriff „commonsense- knowledge- problem“. Man könnte auch von „Hintergrundverständnis“ oder „gesundem Menschenverstand“ sprechen. Auch hierbei handelt es sich um eines der zentralen Dinge, die einem Computer im Vergleich zu uns Menschen fehlen. Um die Frage zu beantworten, warum das Alltagswissen einem Computer nicht einprogrammiert werden kann, müssen wir zunächst die Merkmale desselben betrachten.

Zum einen ist das Hintergrundverständnis implizit. Das heißt, dass uns seine Mitwirkung an kognitiven Operationen ist uns gar nicht bewusst. Ein weiteres Merkmal ist, dass das Alltagswissen holistischer Natur ist. Das bedeutet, es lässt sich nicht einfach als Summe von noch so vielen Elementen oder Regeln darstellen. Entscheidend ist die Anordnung und Verknüpfung der Elemente zu einem netzartigen Ganzen. Weiters ist das Alltagsverständnis ein praktisches Wissen, das durch aktuelle Intentionen, langfristige Interessen und unbewusste Zielsetzungen bestimmt wird. Die Funktion des Alltagswissens liegt in erster Linie darin, uns das alltägliche Handeln in der Welt zu ermöglichen. Schließlich ist das Hintergrundverständnis zum Teil auch leibgebunden, also verkörpertes Wissen. Somit ist der Körper eine wesentliche Bedingung für das Alltagswissen, und bislang ist es noch nicht gelungen, eine Maschine mit einem adäquaten Körper auszustatten. Als letztes Merkmal des gesunden Menschenverstands wollen wir noch nennen, dass er Ausdruck einer sozialen Lebensform ist. Es gibt also bei uns Menschen so etwas wie ein kollektives Unterbewusstsein. Man versteht sich mit manchen Menschen „blind“, und muss bestimmte Dinge gar nicht aussprechen um sie dem Gegenüber verständlich zu machen.

Aus diesen Merkmalen des Alltagswissens folgt nun, dass man es einem Computer nie einprogrammieren können wird. Das Problem wäre gar nicht, dass der Computer es nicht besitzt, sondern dass der Programmierer über sein eigenes Hintergrundwissen nicht vollständig verfügen kann und somit kein dementsprechendes Programm schreiben kann. (vgl. <http://www.phf.uni-rostock.de/fkw/iph/thies/KI.html>)

## 6.5 Die Mikrowelten

Es hat jedoch eine Sorte von Programmen gegeben, die das Relevanz- Problem und das Frame- Problem umgehen konnten, und auch ohne Alltagswissen auskamen: Die Mikrowelten. Wir wollen nun im Folgenden ein Beispiel- Programm für Mikrowelten vorstellen, das von Terry Winograd entwickelt wurde.

Terry Winograd wurde 1946 in Maryland geboren, studierte Informatik und Linguistik, und promovierte 1970 am MIT. Er war Professor für Informatik an der Stanford Universität, und sein Forschungsschwerpunkt war die Mensch- Maschine- Interaktion. Später wurde seine Arbeit in Computerwissenschaft stark von Dreyfus beeinflusst, und er integrierte schließlich Heidegger in seinen Ansatz, und wandte sich vom klassischen Ansatz ab.

Das Programm SHRDLU, welches er schrieb, ist ein Beispiel für eine Mikrowelt, das auch tatsächlich funktionierte- aber natürlich nur innerhalb der Mikrowelt. Es handelt sich dabei um ein Programm, das natürliche Sprachen verstehen soll. Das ganze ist als „Klötzchenwelt“ organisiert, das heißt es gibt einige geometrische Figuren, mit denen das Programm arbeiten kann. Es kann Tatsachen über die Klötze wiedergeben oder auch lernen, oder einzelne Klötze aufheben und an einen anderen Platz stellen. SHRDLU wurde seinerzeit als bedeutender Fortschritt beim Verstehen natürlicher Sprachen angesehen, doch auch hier kritisiert Dreyfus wieder, dass es sich nicht um richtiger Verstehen im Sinne von „begreifen“ handelt.

Unter folgendem Link ist eine noch lauffähige Version von SHRDLU zu finden:

<http://www.semaphorecorp.com/misc/shrdlu.html>



## 7 Abschließende Bemerkungen

Die Vorstellung einer künstlichen Intelligenz hat den Menschen fasziniert und gleichzeitig auch provoziert, seit er sich bemüht, sein eigenes Wesen zu erforschen.

Jedoch sind die spektakulären Erfolge, die prognostiziert worden waren, nicht eingetroffen. Es gibt noch einige Dinge, die unerreichbar geblieben sind und es wahrscheinlich auch immer bleiben werden, wie z.B. die Nachbildung unseres Alltagswissens, die Konstruktion einer logischen Sprache mit dem Anspruch universeller Wissensrepräsentation und die Kommunikation in natürlicher Sprache.

Man sieht nun, dass sich die KI nicht zu einer Schlüsseltechnologie entwickelt hat, sondern eher zu einer Nischentechnologie, die ihre Anwendung z.B. bei Expertensystemen findet. Es bleibt also das ernüchternde Fazit zu ziehen, dass die KI zwar, nicht zuletzt wegen ihres stark interdisziplinären Charakters, ihren Platz als Bindeglied zwischen Kognitionswissenschaft und Wissenstechnologie in Forschung, Lehre und Anwendung gefunden hat, aber bei weitem nicht an der herausragenden Position, die sie sich selbst in den kühnen Träumen ihrer Anfangsjahre zugestehen wollte.

## 8 Literaturverzeichnis

Dreyfus, H. L. (1985). *Die Grenzen künstlicher Intelligenz : was Computer nicht können*.  
Königstein: Athenäum

Dreyfus, H. L., *Why Heideggerian AI failed and how fixing it would require making it more Heideggerian*“.

<http://www.wikipedia.de>

<http://www.ifi.unizh.ch/groups/ailab/teaching/semi2000/Folien.pdf>

<http://www.cse.unsw.edu.au/~achim/Research/Philosophie/node62.html> - 8k

<http://www.ikw.uni-osnabrueck.de/~roul/3%20%20Schritte%20zur%20Komparativen%20Kognitiven%20Robotik.pdf>

<http://www.phf.uni-rostock.de/fkw/iph/thies/KI.html>

<http://www.semaphorecorp.com/misc/shrdlu.html>