

## Modellierung mit Hilfe wissensbasierter Systeme

***Klaus Opwis und Hans Spada***

Darum ist ja die künstliche Intelligenzforschung für die kognitive Psychologie so interessant: Ihre Programme bilden Handlungssysteme . . . so exakt und in so durchsichtiger Weise ab, daß sie Theorien gleichen, die man laufen lassen kann.

(Aebli, 1980)

### ***1. Prolog: Von Brücken aus Bausteinen, Ökosystemen in Form von Differentialgleichungen und Problemlöseprozessen auf Rechnern***

Ein Kind erzählt von einer Brücke. Sie sei groß und verbinde die Ufer eines Flusses; sie glänze in der Sonne und wenn schwere Lastwagen über sie führen, dann bebe sie. Daheim, in seinem Spielzimmer, beginnt das Kind, eine Brücke zu bauen. Es hat einen Baukasten mit vielen verschiedenen Elementen, so daß es in seinem kleinen Modell Pfeiler nachbauen kann, tragende Balken verwenden, Verzierungen anbringen usw. Es ist ganz in sein Tun vertieft, denn es macht Spaß, Brücken zu konstruieren und sei es auch nur aus Bausteinen. Ohne Zweifel hat das Kind sein Wissen über Brücken und über diese eine, besonders faszinierende, sowohl in der verbalen Kommunikation wie auch in der Rekonstruktion mit Elementen des Baukastens zum Ausdruck gebracht. Im letzteren Fall schafft es sich eine Welt, mit der es experimentieren kann; so prüft es beispielsweise, unter welchen Bedingungen sein Bauwerk zusammenbricht. Falls das Kind tatsächlich versucht, die gesehene Brücke nachzubauen, wird der Erfolg umso größer sein, je besser es sich die Brücke eingepägt hat, je reichhaltiger der Baukasten mit verschiedenen Bestandteilen ausgestattet ist und je kompetenter das Kind damit umgehen kann. Natürlich bleibt die Rekonstruktion ein Modell, zugleich ist es aber ein Beispiel für eine Brücke. Das Modell ist eine Brücke.

Ein Biologe untersucht die Belastung eines Binnensees durch Abwässer. Er ist insbesondere an Fragen der Eutrophierung interessiert. Umfassende Da-

tenreihen zu wichtigen biotischen und abiotischen Variablen werden erhoben. Der Biologe spezifiziert ein System von Differentialgleichungen, das die Abhängigkeit der verschiedenen Komponenten abbildet und damit Aspekte der Veränderung des Ökosystems See über die Zeit zu simulieren gestattet. Prognosen über den zukünftigen Entwicklungsverlauf werden möglich. Modell-daten können auch in dem Sinne experimentell generiert werden, als unter Voraussetzung unterschiedlicher Annahmen über die zukünftige Abwasserbelastung bestimmte Eutrophierungsverläufe resultieren. Die Prüfung der empirischen Gültigkeit des Modells erfolgt über den Vergleich simulierter Daten mit (weiteren) Beobachtungsdaten, aber auch anhand verschiedener Plausibilitätsprüfungen. Wichtige Voraussetzungen für ein valides Modell sind vor allem die Kompetenz des Biologen in Fragen seines Gegenstandsbereichs und der mathematischen Systemanalyse, die Qualität und der Umfang der erhobenen Daten, die in die Modellierung eingehen, und die Differenziertheit, mit der etwa die Dynamik zeitkontinuierlich ablaufender Prozesse im verwendeten mathematischen Kalkül abgebildet werden kann. Allerdings ist die Simulation eines Binnensees mit Hilfe mathematisch-numerischer Modelle nicht selbst ein Binnensee, da das Simulierte und seine Simulation in diesem Fall auf unterschiedlichen Prinzipien beruhen.

Eine Denkpsychologin beobachtet Probanden bei der Bearbeitung des ‚Turm von Hanoi‘-Problems. Sie stellt fest und berichtet dies auch in ihrer Forschungsgruppe, daß die Schwierigkeit der Problemlösung mit der Anzahl der Scheiben zunehme, und zwar insbesondere bei ungetübten Personen. Viele fehlerhafte Züge seien zu beobachten, aber sie habe schon eine Vermutung, in welchen Fällen das Erkennen einer Sackgasse von zentraler Bedeutung für die Aufdeckung des optimalen Vorgehens sei. An ihrem Schreibtisch beginnt die Psychologin, den Problemlöseprozeß, der dem beobachteten Verhalten zugrundeliegen könnte, auf dem Rechner zu rekonstruieren. Sie wählt eine symbolische Repräsentation für das Problem, spezifiziert auf dieser Grundlage das Vorwissen der simulierten Probanden und entwickelt einen Lösungs- und Lernalgorithmus, der in einer geeigneten Programmiersprache als wissensbasiertes System auf dem Rechner implementiert wird. Das Ergebnis ist eine Modellierung, für die die Wissenschaftlerin prüft, ob sie das beobachtete Lösungsverhalten in wichtigen Charakteristika korrekt generiert. Darüber hinaus testet sie weitere, aus der Modellierung abgeleitete Annahmen anhand zusätzlicher empirischer Untersuchungen. Die Modellierung wird um so valider sein, je detaillierter die Information ist, die empirisch über diesen kognitiven Prozeß vorliegt. Wichtig sind die Adäquatheit und Reichhaltigkeit der formalen Sprache, mit deren Hilfe die Modellierung erfolgt, aber auch die Kompetenz der Psychologin im Umgang mit den Elementen dieser Sprache. Natürlich bleibt die Rekonstruktion ein Modell, aber sie ist zugleich ein Beispiel für einen Problemlösungsprozeß.

Können zwischen dem Nachbauen einer Brücke mit den Bestandteilen eines Baukastens, der Modellierung der Eutrophierung eines Sees anhand von Differentialgleichungssystemen und der Rekonstruktion eines Problemlöseprozesses mit Hilfe eines wissensbasierten Systems Analogien gezogen werden?

Allen drei Beispielen ist gemeinsam, daß eine ausschließlich natürlichsprachliche Erfassung des Gegenstandsbereiches, sei es in der Umgangssprache oder einer Fachsprache, zu kurz greift. Erst das Modell aus Bausteinen oder aber die formalisierten Prozeßmodelle erlauben einen aktiven Umgang mit der Modellierung. Bedingungsmanipulationen und Prognosen werden möglich.

Das Brückenbaubeispiel und den Fall der auf dem Rechner simulierten Problemlösungsprozesse verbinden in der Phase der Modellerstellung das ingenieurmäßige, konstruktive Vorgehen in der Modellierung und der geringe Grad an Datenrestriktion (vgl. Tack, in diesem Band). Es handelt sich um ein eher interpretatives Vorgehen. Während keine Parallelen zwischen einem kindlichen Vorgehen und denkpsychologischen Forschungsstrategien gezogen werden sollen, ist der Bezug zwischen ‚Das Modell einer Brücke ist eine Brücke‘ und ‚Die rekonstruktive Modellierung einer Problemlösung ist eine Problemlösung‘ durchaus intendiert. Für die Ökosystemmodellierung und die denkpsychologische Simulation gilt, daß es sich in beiden Fällen um formale Modelle handelt, allerdings im ersten Fall um ein mathematisch-numerisches Modell, im zweiten Fall um ein wissensbasiertes Modell. Beides sind Prozeßmodelle. Eine hohe Datenrestriktion charakterisiert die Phase der Modellgeltungsprüfung.

Halten wir fest: Geistige Prozesse, die uns in der Reflexion bewußt werden, die wir aber auch als Erklärung für das Verhalten anderer heranziehen, lassen sich in unterschiedlicher Weise, umgangssprachlich, fachsprachlich oder formal beschreiben. Wir können Aussagen über ihre Funktion formulieren, über ihr Ziel und über die Bedingungen, unter denen sie auftreten. Darüber hinaus können wir sie aber auch modellieren, indem wir sie formal fassen und die Formalisierung auf einem Rechner in einer geeigneten Programmiersprache implementieren. Man kann dabei Prozesse abbilden, die zu korrekten, aber auch andere, die zu fehlerhaften Ergebnissen führen, man kann eine bestimmte Person in einer definierten Situation vor Augen haben oder aber kognitive Prozesse in prototypischer Weise abzubilden versuchen. Unter dem Blickwinkel der Psychologie bedarf die Modellierung auf dem Rechner der Validierung anhand empirischer Daten. Ziel ist nicht nur eine Rekonstruktion, die bei gegebenen Eingaben zum selben Ergebnis führt, das beim Menschen beobachtet wurde, sondern auch die Nachbildung wesentlicher Ablaufcharakteristika.

## 2. Modellierung als Rekonstruktion

Das Denken und Handeln von Personen beruht auf ihren Deutungen und Auslegungen der sie umgebenden Welt. Personen konstruieren durch Vorgänge der Wahrnehmung und des Sprachverstehens sowie aufgrund von (Vor-) Wissen, Erwartungen und Zielsetzungen eine interne, subjektive Repräsentation von Ausschnitten ihrer tatsächlichen oder einer vorgestellten Umgebung. Diese **Konstruktionen erster Ordnung** oder **mentalen Modelle'** sind als Rekonstruktion der Umgebung wie jede Art der Modellbildung eine Abstraktion. Dies impliziert im Besonderen ihre prinzipielle Unvollständigkeit. Darüber hinaus sind sie nicht unmittelbar beobachtbar. Der Beobachtung zugänglich sind lediglich Verhaltensdaten, aus denen auf das zugrundeliegende mentale Modell geschlossen werden muß. Damit wird die psychologisch valide Rekonstruktion des mentalen Modells einer Person zur zentralen Zielsetzung der psychologischen Theoriebildung. Die in wissenschaftlicher Perspektive zu erarbeitenden **Modelle mentaler Modelle** sind in dieser Hinsicht **Konstruktionen zweiter Ordnung**.

Die Thematik dieses Kapitels ist eine spezielle Methodik zur Rekonstruktion mentaler Modelle, ihre **Modellierung mit Hilfe wissensbasierter Systeme**. Die theoretische Rahmenvorstellung ist das Paradigma der Informationsverarbeitung, das die kognitiv orientierte Psychologie seit Ende der sechziger Jahre dominiert. Grundannahme ist, daß geistige Aktivitäten wie induktive und deduktive Denkvorgänge, Problemlösen, Planungsprozesse, Sprechen und Sprachverstehen, aber auch die Wahrnehmung der Umwelt, als **wissensbasierte Vorgänge der Symbolverarbeitung** aufgefaßt werden können. Diese Betrachtungsweise wird in den weiteren Ausführungen auf der Grundlage des Konzepts wissensbasierter Systeme elaboriert. Ziel ist die Explikation zentraler methodologischer Annahmen und Argumente, die der Modellierung mentaler Modelle durch wissensbasierte Systeme zugrundeliegen.

Damit wird in diesem Kapitel die Auffassung vertreten, daß die Analyse geistiger Aktivitäten als symbolische Berechnungsvorgänge und ihre korrespondierende Rekonstruktion mit technischen Mitteln aufgrund der dazu erforder-

---

1 Der Begriff ‚mentales Modell‘ wird in der kognitiven Psychologie keineswegs einheitlich verwendet (vgl. auch Herrmann, 1988). In der vorliegenden Arbeit ist damit ein Ausschnitt der internen Repräsentation der äußeren Welt eines Subjekts gemeint, die es ihm ermöglicht, Phänomene zu verstehen, Vorhersagen zu treffen und Inferenzen durchzuführen. Ein wesentliches Charakteristikum ist die Möglichkeit von ‚inneren Simulationen‘, etwa in Form von Schlußfolgerungen und mentaler Probehandlungen, über die An der Repräsentation werden keine Annahmen getroffen. Demgegenüber betrachtet etwa Johnson-Laird (1983) mentale Modelle als eine spezifische Form der Wissensrepräsentation, der sog. analogen Repräsentation (in Abgrenzung von propositionalen und regelbasierten Repräsentationssystemen). Kennzeichen ist bei ihm eine erfahrungsmäßige Komponente, einhergehend mit Vorstellungsbildern, in denen räumliche Relationen zwischen Objekten und/oder zeitliche Relationen zwischen Ereignissen repräsentiert sind,

derlichen lückenlosen Formalisierung und kompromißlosen Explizitheit besonders geeignet sind, zentrale inhaltliche und methodologische Fragen offenzulegen. Zusammenfassend wird eine Position eingenommen, wie sie etwa von Johnson-Laird (1988) formuliert wurde.

Is the mind a computational phenomenon? No one knows. . . . what is clear is that **computability provides an appropriate conceptual apparatus for theories of the mind** (Hervorhebung von den Verfassern). This apparatus takes nothing for granted that is not obvious . . . Yet, . . . any clear and explicit account of, say, how people recognize faces, reason deductively, create new ideas or control skilled actions can always be modelled by a computer program.

(Johnson-Laird, 1988, S. 51f.)

Danach sind Computer nützliche analytische Werkzeuge, um unser Verständnis kognitiver Vorgänge beim Menschen zu verbessern, indem - wie so oft in der Wissenschaftsgeschichte - eine neuartige, aber bekannte Technologie zur Grundlage unseres Verstehens des Unbekannten gemacht wird (vgl. Gigerenzer, 1988, 1991). Computermodelle erlauben die Formulierung und Prüfung theoretischer Vorstellungen in einer Form, die durch die grundlegenden wissenschaftstheoretischen Ideale der **Präzision, Widerspruchsfreiheit** und **Vorhersagbarkeit** gekennzeichnet ist. Die These ist, daß die Psychologie mit den Möglichkeiten der Modellierung geistiger Prozesse am Rechner ein Werkzeug in die Hand bekommen hat, das mächtiger ist als die bisher verfügbaren Ansätze der Modellbildung.

Da in diesem Kapitel methodologische Fragen im Vordergrund der Betrachtung stehen, soll zumindestens über Literaturhinweise auch ein kurzer Überblick über Inhalte kognitiver Modellierungen gegeben werden.

Eine historisch orientierte, breit angelegte Darstellung der Kognitionswissenschaft enthält Gardner (1985). Eine Sammlung wichtiger ‚klassischer‘ Arbeiten findet man in Collins und Smith (1988). Das Buch von Johnson-Laird (1988) stellt ein einführendes, an Inhalte der Allgemeinen Psychologie anknüpfendes Lehrbuch dar, während die von Posner (1989) bzw. Osherson (Osherson & Lasnik, 1990; Osherson, Kosslyn & Hollerbach, 1990; Osherson & Smith, 1990) herausgegebenen Bände handbuchartig in die Kognitionswissenschaft einführen. Ein Überblick zur Rolle der kognitiven Modellierung in der psychologischen Theoriebildung findet sich bei Opwis (1992).

Es folgt eine Auflistung grundlegender klassischer, aber auch beispielhafter moderner Arbeiten zu ausgewählten inhaltlichen Themenbereichen. Zum **Problemlösen und Lernen** sind insbesondere die Arbeiten von Newell und Simon sowie von Anderson zu nennen (etwa: Newell, Shaw & Simon, 1958; Newell & Simon, 1963, 1972; Anzai & Simon, 1979; Anderson, 1982, 1983, 1987). Daneben sei auf die von Holland, Holyoak, Nisbett und Thagard (1986) vorgeschlagene Theorie induktiven Denkens und Lernens sowie auf die Arbeiten

von VanLehn zum Erwerb einfacher prozeduraler Fertigkeiten und zur Rolle des Verstehens von Beispielen für den Erwerb von Problemlöseprozeduren und deren Transfer auf neue Probleme hingewiesen (vgl. VanLehn, 1989; VanLehn, Jones & Chi, 1992). Vielzitierte Autoren zur Wissensrepräsentation sind Quillian (1968) sowie Collins und Loftus (1975) zu semantischen Netzen, Minsky (1975) zu „Frames“ oder Schank und Abelson (1977) zum Skriptkonzept. Eine Sammlung grundlegender Arbeiten findet sich in Brachman und Levesque (1985), ein modernes Lehrbuch ist Davis (1990). Aus der Fülle von einschlägigen Arbeiten zur **Sprache** seien Chomsky (1963), Winograd (1972, 1983) und Allen (1987) herausgegriffen. Berwick (1985) und Pinker (1989) behandeln das Thema Spracherwerb. Informativ orientierte Arbeiten zur Wahrnehmung sind etwa Marr (1982) und Hildreth und Ullman (1989). Schließlich seien noch die Veröffentlichungen von Anderson (1983) und Newell (1990; vgl. auch Laird, Newell & Rosenbloom, 1987) zur Frage der **Architektur des menschlichen kognitiven Apparats** genannt.

Viele wichtige kognitionswissenschaftliche Originalarbeiten werden in der Zeitschrift **Cognitive Science** veröffentlicht. Seit 1991 gibt es auch eine deutschsprachige Zeitschrift **Kognitionswissenschaft**. Daneben sind eine Reihe weiterer Fachzeitschriften zu nennen, wie das **Journal of the Learning Sciences**, **User modeling and User-adapted Interaction**, **Minds and Machines**, sowie mit einer stärkeren KI-Orientierung etwa **Artificial Intelligence**, **Machine Learning** oder **Journal of Artificial Intelligence in Education**. Darüber hinaus veranstaltet die Cognitive Science Society jährliche Konferenzen, deren Tagungsberichte als **Proceedings of the Conference of the Cognitive Science Society** veröffentlicht werden.

Am Anfang der hier angesprochenen Entwicklung der Kognitionswissenschaft, etwa Mitte der fünfziger Jahre, haben Psychologen und Informatiker Hand in Hand gearbeitet. Etwa seit Mitte der siebziger Jahre dominiert zunehmend der Beitrag der Informatik, zumindestens hinsichtlich der Weiterentwicklung der begrifflichen und methodischen Grundlagen. Die Psychologie muß alle Anstrengungen unternehmen, damit sie auch künftig einen gestaltenden Anteil an einer multidisziplinären kognitiven Wissenschaft hat.

Die weiteren Inhalte und der Aufbau des Beitrags lassen sich wie folgt zusammenfassen. Im dritten Abschnitt folgt eine Darstellung ausgewählter theoretischer und formaler Grundlagen. Damit soll eine Verständigung über die Grundbegriffe und die Untersuchungsperspektive erreicht werden. Anhand eines Beispiels wird das Problem der Rekonstruktion menschlichen Wissens mit technischen Mitteln illustriert. Dies führt zu einer Explikation des Konzepts wissensbasierter Systeme und zur **Repräsentationshypothese**. Danach sind wissensbasierte Systeme algorithmische Modelle, die durch die Verwendung und Manipulation formaler Symbolstrukturen gekennzeichnet sind. Es

folgt eine Präzisierung des **Algorithmenbegriffs** und des damit aufs engste verbundenen Konzepts der **Berechenbarkeit**. Danach steht die Verwendung wissensbasierter Systeme als kognitionspsychologische Modelle im Mittelpunkt. Das Verhalten, die Funktionalität und die Struktur technischer, aber auch biologischer Systeme lassen sich auf unterschiedlichen Betrachtungsebenen beschreiben, analysieren und erklären. Dazu wird eine psychologisch motivierte **Mehrebenenbetrachtung** informationsverarbeitender Systeme entwickelt, die einen konstruktiven Umgang mit dem in diesem Zusammenhang zentralen Problem der Vermischung von Ebenen ermöglicht.

Fragen der Bewertung wissensbasierter Systeme als psychologische Modelle werden im vierten Abschnitt behandelt. Dabei wird insbesondere das Kriterium der empirischen Adäquatheit betrachtet und es werden die verschiedenen Funktionen diskutiert, die empirische Befunde im Prozeß der Theorieformulierung und -prüfung erfüllen.

Einige wissenschaftstheoretische Anmerkungen und ein Epilog runden das Kapitel ab.

### 3. Grundlagen wissensbasierter Systeme

Nach der allgemeinen Modelltheorie (Stachowiak, 1973) sind **Modelle** immer Modelle von **etwas**, Abbildungen, Repräsentationen bzw. Rekonstruktionen natürlicher oder künstlicher Originale, die ihrerseits auch wieder Modelle sein können (vgl. hierzu auch den Modellbegriff der (Prädikaten-) Logik; z.B. Genesereth & Nilsson, 1987). Modelle sind dabei ihren Originalen nicht per se zugeordnet, sondern sie erfüllen ihre Funktionen stets für bestimmte Erkenntnissubjekte und/oder Akteure und relativ zu bestimmten Zielen und Zwecken. Hinsichtlich des Mediums, in dem die Modellbildung vorgenommen wird, lassen sich zwei große Klassen von Modellen unterscheiden (vgl. Gigerenzer, 1981): **Symbolische** und **analoge** Modelle.

Bei **symbolischen Modellen** erfolgt die Modellbildung in einem Medium, das - meist per Definition - in eine Menge von elementaren Informationseinheiten (Symbole) zerlegbar ist und über eine Menge von Vorschriften zur Zusammensetzung von komplexen Symbolstrukturen verfügt. Dies impliziert die Eigenschaft der **Grammatikalität** bzw. eine Definition von **syntaktischer Wohlgeformtheit**. Die wichtigsten Unterklassen sind **formale Modelle** numerischer (etwa: probabilistische mathematische Modelle) oder **nicht-numerischer Art** (etwa: logische Kalküle oder mengen- und graphentheoretische Vorstellungen) sowie **natürlich-sprachliche Modelle** (etwa: Theorien auf der Grundlage einer mehr oder weniger präzise definierten **fachwissenschaftlichen Terminologie** oder der **Alltagssprache**).

Hinsichtlich ihrer Brauchbarkeit als wissenschaftliche Modelle weisen formale gegenüber sprachlich formulierten Modellen eine Reihe von Vorteilen auf (Vermeidung von widersprüchlichen Annahmen, Nachweis von lückenhaften Annahmen, Präzisierung der verwendeten Begrifflichkeit). Ausführlich werden die Funktion formaler Modelle, ihre spezifischen Vorteile und Probleme etwa in Deppe (1977) oder Neches (1982) erörtert. Die Unterschiede zwischen numerischen und nicht-numerischen Modellen liegen nicht nur in der Art ihrer Formulierung. Mathematische Modelle auf der Grundlage statistischer oder wahrscheinlichkeitstheoretischer Überlegungen sind Modelle von Daten. Sie zielen zumeist auf ausgewählte und direkt beobachtbare Aspekte psychischer Aktivitäten wie etwa Reaktionszeiten. Sie können in diesem Sinne auch als **„oberflächliche“** Theorien nach Moravcsik (1980) charakterisiert werden. Nicht-numerische Modelle thematisieren dagegen als **„tiefte“** Theorien die dem Verhalten von Personen als zugrundeliegend angenommenen und prinzipiell nicht unmittelbar beobachtbaren psychischen Strukturen und Prozesse selbst. In der Psychologie hatten mathematische Modelle insbesondere in den fünfziger und sechziger Jahren ihre Blütezeit (vgl. das dreibändige Handbuch der mathematischen Psychologie; Luce, Bush & Galanter, 1963), während nicht-numerische Modelle seit den siebziger Jahren dominieren. Allerdings ist in Gestalt des sog. **„Neueren Konnektionismus“** gegenwärtig eine Renaissance einer Klasse numerischer Modelle erkennbar, die Analogien zu neuronalen Prozessen der Informationsverarbeitung aufweisen. Auf die theoretischen Grundlagen dieser, in der Regel als nicht-lineare Differenzen- oder Differentialgleichungssysteme formulierten Modelle kann im Rahmen dieses Beitrags nicht eingegangen werden (vgl. dazu etwa McClelland & Rumelhart, 1986, 1988; Rumelhart & McClelland, 1986; Smolensky, 1988; Strube, 1990).

Die in diesem Beitrag näher ausgeführte Methode zur Rekonstruktion kognitiver Strukturen und Prozesse ist ihre formale, nicht-numerische Modellierung (vgl. Tack, 1987). Das Problem der Modellbildung kann umschrieben werden als das Problem der adäquaten Rekonstruktion „von strukturell-operativen Zusammenhängen menschlichen Wissens in Algorithmen und Datenstrukturen“ (Scheffe, 1986, S.155). Die Besonderheiten und Anforderungen dieser Form der Modellbildung werden im folgenden Abschnitt anhand eines bekannten gedächtnispsychologischen Beispiels, den sog. semantischen Netzwerkmodellen, illustriert. Dabei geht es hier nicht primär um den spezifischen Inhalt dieser Modelle. Vielmehr soll aufgezeigt werden, daß bei der Modellbildung drei unterschiedliche, aber eng aufeinander-bezogene Ebenen relevant sind: Die Ebene der Konzeptualisierung als Formulierung theoretischer Vorstellungen mit Hilfe psychologischer Begriffe, die Ebene der Formalisierung dieser Vorstellungen mit formal-logischen Kalkülen und die Ebene der Implementation als wissensbasierte Systeme.

### 3.1 Konzeptuelle Modelle, formale Kalküle und wissensbasierte Systeme am Beispiel semantischer Netzwerkmodelle

Im Zusammenhang mit der Frage, wie das sprachlich-begriffliche Wissen einer Person im Gedächtnis gespeichert ist, taucht in der psychologischen Literatur immer wieder der Begriff des **semantischen Netzes** auf. Allerdings werden von verschiedenen Autoren sehr unterschiedliche Modellvorstellungen unter diesen Begriff subsumiert (vgl. Opwis & Lüer, in Druck). Gemeinsam ist der theoretische Anspruch, daß sie ein **Modell der Organisation sprachlicher Gedächtnisinhalte** darstellen. In der Mehrzahl der Modelle wird als zugrundeliegende formale Struktur ein **gerichteter Graph** (bzw. allzu oft auch nur eine entsprechende graphische Notation) verwendet. Die Knoten und Kanten eines Graphen stehen für sprachliche Einheiten und Relationen. Die Modelle unterscheiden sich nun insbesondere dadurch, von welchen elementaren sprachlichen Einheiten (insbesondere Begriffe oder Sätze bzw. Propositionen) sie ausgehen, welche klassifikatorischen (etwa Unter-/Oberbegriff), modifikatorischen (Attributzuschreibung), syntaktischen (Nominal-, Verbalphrase usw.) oder semantischen (Handlung, Akteur, Objekt u. ä.) Relationen zugrundegelegt werden und welche Prozesse (Vergleichs-, Inferenzprozesse oder Aktivierungsvorgänge) im Mittelpunkt der Analyse stehen.

#### **3.1.1 Konzeptualisierung: Abstraktionshierarchien als psychologische Modellvorstellung**

Eine der wichtigsten Anwendungen semantischer Netzwerkmodelle ist die Beschreibung von Taxonomien, das sind hierarchisch geordnete Klasseneinteilungen eines Gegenstandsbereichs. Man spricht in diesem Zusammenhang auch von **Abstraktionshierarchien**. Ein bekanntes, in der einschlägigen Literatur seit Anfang der siebziger Jahre immer wieder bemühtes Beispiel ist die folgende, in Abbildung 1 veranschaulichte Abstraktionshierarchie von Tierarten (vgl. Klix, 1988; Möbus, 1988).

Wie man unmittelbar erkennt, ist die Abstraktionshierarchie eine netzartige Struktur, die auf einer Einteilung in Ober- bzw. Unterklassen (bzw. -begriffe) im Sinne einer ‚ist-Teilmenge-von‘ Relation beruht. Beispielsweise gelangt man ausgehend von der Klasse der Lachse in zwei Abstraktionsschritten zur Klasse der Tiere (über Fisch). Die minimale Länge des Weges zwischen zwei Begriffen definiert ihre semantische Distanz. Eine bestimmte Klasse ist nun durch eine Menge von Attributen, Merkmalen oder Eigenschaften definiert. Abstraktion bedeutet Elimination von Merkmalen. Je abstrakter eine Klasse ist, desto weniger charakterisierende Merkmale weist sie auf. Wichtig ist in diesem Zusammenhang aber die Vorstellung **kognitiver (Speicher-) Ökonomie**.

Informationen werden nicht mehrfach abgespeichert, sondern anschaulich ausgedrückt an dem ‚obersten‘ Knoten, für den sie Gültigkeit haben, angehängt und im Bedarfsfall ‚nach unten‘ weitergereicht. Die Abstraktionshierarchie bildet somit zugleich Implikationen und mögliche Inferenzwege ab. Eigenschaften, die auf Oberbegriffe zutreffen, können über diese Wege auf alle ihre Unterbegriffe vererbt werden. Beispielsweise kann eine Frage der Art ‚Hat ein Lachs Kiemen?‘ auf diese Weise positiv beantwortet werden (Da jeder Lachs ein Fisch ist).

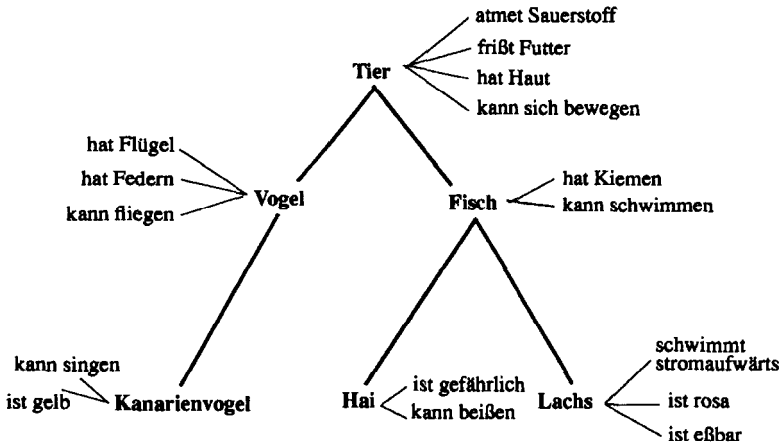


Abb. 1: Taxonomie von Tierarten als Beispiel für eine Abstraktionshierarchie (nach Collins & Quillian, 1969)

Im nächsten Schritt soll die bisher verwendete psychologische Begrifflichkeit präzisiert werden. Dazu ist eine Festlegung der **Objekte** erforderlich, über die Aussagen gemacht werden sollen, sowie die Definition von **Relationen** zwischen den Objekten und von **Operationen** auf den Objekten bzw. Relationen. Dieser Vorgang wird auch als **Konzeptualisierung** bezeichnet (vgl. Genesereth & Nilsson, 1987).

Für dieses Beispiel liegt es nahe, die relevanten Objekte mit den Knoten bzw. mit den in der Abstraktionshierarchie vorkommenden Begriffen zu identifizieren:

{Tier, atmet-Sauerstoff, . . . , Vogel, ...}

Die Menge der in einer Konzeptualisierung betrachteten Objekte definiert zugleich ihren sog. Diskursbereich. Eine Relation ist eine spezielle Beziehung zwischen Objekten des Diskursbereichs. Beispielsweise bilden die Kanten in Abbildung 1 zwei unterschiedlichen Arten von Relationen ab, eine klassifikatorische Relation im Sinne der Unter-/Oberbegriffsrelation und eine modi-

fikatorische Relation, die einem Begriff eine spezielle Eigenschaft zuordnet. Damit korrespondieren die beiden folgenden Mengen von Tupeln:

$$\{ \langle \text{Vogel}, \text{Tier} \rangle, \langle \text{Fisch}, \text{Tier} \rangle, \dots \}$$

und  $\{ \langle \text{Tier}, \text{atmet\_Sauerstoff} \rangle, \langle \text{Tier}, \text{frisst\_Futter} \rangle, \dots \}$

Diese beiden Relationen stellen keineswegs die einzig möglichen Relationen zwischen den betrachteten Begriffen dar. Eine Konzeptualisierung, die aus einer bestimmten Menge von Objekten und einer Menge von betrachteten Relationen besteht, ist prinzipiell unvollständig. Unter Umständen verhindert die gewählte Konzeptualisierung die Darstellung bestimmter Sachverhalte. Ein bekanntes Beispiel hierfür ist die Kontroverse in der Physik über die Natur von Licht. Je nachdem, ob Licht als ein Teilchen- oder als ein Wellenphänomen konzeptualisiert wird, können spezielle Aspekte des Verhaltens von Licht erklärt werden, aber keine der Sichtweisen reicht allein aus. Erst die Entwicklung einer weiteren Konzeptualisierung auf der Grundlage der modernen Quantentheorie hat diese Schwierigkeiten beseitigt, indem sie beide Sichtweisen zusammenführte.

### **3.1.2 Formalisierung: Prädikatenlogische Darstellung von Abstraktionshierarchien**

Haben wir uns für eine Konzeptualisierung entschieden, so können wir die Formalisierung des Wissens vornehmen. Dazu wollen wir im folgenden Abschnitt eine formale Sprache verwenden, das sog. Prädikatenkalkül bzw. die Prädikatenlogik. Die wesentlichen Vorteile sind:

- die Unterscheidung von Variablen und Konstanten, wobei Variablen benötigt werden, um Eigenschaften von Objekten darzustellen, ohne dabei diese Objekte explizit benennen zu müssen;
- die Verwendung logischer Operatoren wie  $\wedge$  (Konjunktion),  $\vee$  (Disjunktion) und  $\Rightarrow$  (Implikation), die es ermöglichen, aus einfachen komplexe Aussagen zu bilden, ohne dabei die Wahrheit oder Falschheit der einfachen Aussagen angeben zu müssen;
- die Verwendung von Quantoren wie  $\forall$  (Allquantor) und  $\exists$  (Existenzquantor), die es ermöglichen, über alle Objekte Aussagen zu machen, ohne die Objekte im einzelnen aufzählen zu müssen bzw. die Existenz eines Objektes mit bestimmten Eigenschaften anzunehmen, ohne das Objekt selbst identifizieren zu müssen;
- eine präzise Definition der Bedeutung prädikatenlogischer Ausdrücke (denotative Semantik).

Zur Formalisierung unseres Beispiels benötigen wir nicht das gesamte Sprachrepertoire der Prädikatenlogik, sondern nur ausgewählte Bestandteile. Dabei werden Konstanten bzw. Variablen im folgenden durch eine Folge aus Buchstaben oder Ziffern dargestellt, wobei das erste Zeichen bei einer Konstanten ein Großbuchstabe und bei einer Variablen ein Kleinbuchstabe ist. Die Tatsache, daß der Begriff ‚Vogel‘ (als Objektkonstante) ein Unterbegriff des Begriffs ‚Tier‘ ist, kann dann durch die Einführung eines konstanten Relationensymbols ‚Ist-ein‘ in der folgenden Weise notiert werden:

Ist\_ein(Vogel,Tier)  
Ist\_ein(Fisch,Tier) usw.

Entsprechend kann die zweite Relationsart durch die Einführung eines weiteren Relationensymbols ‚Hat\_Eigenschaft‘ notiert werden:

Hat\_Eigenschaft(Tier,Atmet\_Sauerstoff)  
Hat\_Eigenschaft(Tier,Frißt\_Futter) usw.

Auf diese Weise können die betrachteten Relationen zwischen Begriffen als eine Menge von **Propositionen** im Sinne logischer Axiome formalisiert werden.

Allerdings sind mit dieser Formalisierung noch keine Inferenzvorgänge spezifiziert. Fragen der Art ‚Ist ein Lachs ein Tier?‘ oder ‚Hat ein Lachs Kiemen?‘ können nicht beantwortet werden. Dazu sind Operationen im Sinne logischer Implikationen erforderlich, die Beziehungen zwischen Relationen definieren. Sie legen fest, wie aus gegebenen Fakten (Axiome oder Prämissen) neue Aussagen (Theoreme oder Konklusionen) abgeleitet werden können. Die Ober-/Unterbegriffsrelation zwischen zwei beliebigen Begriffen einer Abstraktionshierarchie läßt sich rekursiv mit Hilfe zweier Regeln **O1** (Terminierungsfall) und **O2** (rekursiver Fall) definieren:

- (O1) begriff1 ist ein Unterbegriff von begriff2, wenn zwischen beiden Begriffen eine ‚Ist-ein‘ Relation besteht.
- (O2) begriff1 ist ein Unterbegriff von begriff2, wenn zwischen begriff1 und einem begriff3 eine ‚Ist-ein‘ Relation besteht und begriff3 ein Unterbegriff von begriff2 ist.

Die prädikatenlogische Formalisierung dieser Regeln ist:

$\text{Qx Qy} \quad \text{Ist\_ein}(x,y) \Rightarrow \text{Ist\_Unterbegriff}(x,y)$

$\text{Qx Qy} \quad \forall z \text{ Ist\_ein}(x,z) \wedge \text{Ist\_Unterbegriff}(z,y) \Rightarrow \text{Ist\_Unterbegriff}(x,y)$

Zur vollständigen Konzeptualisierung der Abstraktionshierarchie fehlen noch Inferenzregeln zur Vererbung von Eigenschaften. Dazu sind ebenfalls zwei Regeln erforderlich:

- (O3) begriff1 besitzt die Eigenschaft x, wenn zwischen begriff1 und x eine ‚Hat\_Eigenschaft‘ Relation besteht.
- (O4) begriff1 besitzt die Eigenschaft x, wenn begriff1 ein Unterbegriff von begriff2 ist und zwischen begriff2 und x eine ‚Hat\_Eigenschaft‘ Relation besteht.

Die korrespondierende prädikatenlogische Formalisierung dieser Regeln ist:

$$\forall x \forall y \quad \text{Hat\_Eigenschaft}(x,y) \Rightarrow \text{Besitzt\_Eigenschaft}(x,y)$$

$$\forall x \forall y \forall z \quad \text{Ist\_Unterbegriff}(x,z) \wedge \text{Hat\_Eigenschaft}(z,y) \Rightarrow \text{Besitzt\_Eigenschaft}(x,y)$$

Die vorgenommene logikorientierte Konzeptualisierung einer Domäne kann nun im nächsten Schritt als ein wissensbasiertes System implementiert werden. Im Beispiel bietet sich dazu die Programmiersprache PROLOG an (vgl. Sterling & Shapiro, 1986; Malpas, 1987).

### 3.1.3 Implementation: Abstraktionshierarchien als wissensbasierte Systeme

Ein PROLOG-Programm besteht im wesentlichen aus einer Wissensbasis, die alle bekannten Fakten (Axiome) und Regeln (Implikationen) umfaßt. Dabei sind folgende syntaktischen Unterschiede zur logischen Formalisierung im vorigen Abschnitt zu beachten: Regeln und Fakten werden mit einem Punkt abgeschlossen, ihre Namen beginnen ebenso wie konstante Argumente mit einem Kleinbuchstaben, während Variablen mit einem Großbuchstaben beginnen; die konjunktive Verknüpfung von Teilzielen wird durch ein Komma dargestellt. Tabelle 1 enthält die vollständige Implementation einer Abstraktionshierarchie als PROLOG-Programm.

Tabelle 1: Implementation einer Abstraktionshierarchie als PROLOG-Programm

---

<b>% Relationen der Abstraktionshierarchie als PROLOG-Fakten</b>	
ist_ein(vogel, tier).	ist_ein(fisch, tier).
ist_ein(kanarienvogel, vogel).	ist_ein(hai, fisch).
ist_ein(lachs, fisch).	
hat_Eigenschaft(tier, atmet-Sauerstoff).	hat_Eigenschaft(tier, frißt-Futter).
hat_Eigenschaft(tier, hat_Haut).	hat_Eigenschaft(tier, kann_sich-bewegen).
hat_Eigenschaft(vogel, hat_Flügel).	hat_Eigenschaft(vogel, hat_Federn).
hat_Eigenschaft(vogel, kann_fliegen).	hat_Eigenschaft(fisch, hat_Kiemen).
hat_Eigenschaft(fisch, kann_schwimmen).	hat_Eigenschaft(kanarienvogel, ist_gelb).
hat_Eigenschaft(kanarienvogel, kann_singen).	hat_Eigenschaft(hai, ist_gefährlich).
hat_Eigenschaft(hai, kann_beißen).	hat_Eigenschaft(lachs, ist_eßbar).
hat_Eigenschaft(lachs, schwimmt_stromaufwärts)	
hat_Eigenschaft(lachs, ist_rosa).	

---

Tabelle 1 (Fortsetzung)

---

<b>% Regeln zur Definition der Unter-/Oberbegriffsrelation zwischen zwei Begriffen</b>	
ist_Unterbegriff(Unterbegriff, Oberbegriff) :-	ist_ein(Unterbegriff, Oberbegriff)
ist_Unterbegriff(Unterbegriff, Oberbegriff) :-	ist_ein(Unterbegriff, Begriff), ist_Unterbegriff(Begriff, Oberbegriff)
% Regeln zur Vererbung von Eigenschaften	
besitzt_Eigenschaft(Begriff, Eigenschaft) :-	hat_Eigenschaft(Begriff, Eigenschaft).
besitzt_Eigenschaft(Begriff, Eigenschaft) :-	hat_Eigenschaft(Oberbegriff, Eigenschaft).

---

Berechnungen werden in PROLOG als deduktive Berechnungsvorgänge aufgefaßt. Dazu ist neben der Wissensbasis ein domänenunabhängiger, allgemeiner Inferenzmechanismus erforderlich, der auf der Menge der Fakten und Regeln im Sinne eines verallgemeinerten Modus Ponens operiert und durch eine Anfrage angestoßen wird, etwa:

```
?- ist_ein(Objekt,tier).
```

In dieser Anfrage ist ‚Objekt‘ eine (PROLOG-) Variable. Das PROLOG-System prüft, ob die Anfrage aus der Wissensbasis ableitbar ist. Dies ist möglich, wenn die Variable instantiiert werden kann, d.h. wenn in der Wissensbasis eine Konstante gefunden wird, für die sie im Sinne eines Platzhalters stehen kann. Aus Tabelle 1 ist ersichtlich, daß die Variable ‚Objekt‘ durch die Konstante ‚vogel‘ instantiierbar ist. Durch Eingabe eines Semikolons kann eine Anfrage fortgesetzt werden, um evtl. weitere Antworten zu erhalten.

```
?- ist_ein(Objekt,tier).
```

```
Objekt = vogel;  
Objekt = fisch;  
no
```

Beispiele für weitere Anfragen sind etwa für „Ist ein Hai ein Tier?“

```
?- ist_Unterbegriff(hai,tier).
```

oder

```
?- ist_Unterbegriff(Begriff,Klasse).
```

Die letzte Anfrage liefert alle in der Wissensbasis direkt oder indirekt vorhandenen Unter-/Oberbegriffsrelationen. Die nächste Anfrage liefert alle Eigenschaften der Klasse der Kanarienvögel als Antwort, wobei auch die bei den Oberbegriffen nicht redundant gespeicherten Eigenschaften inferiert werden:

```
?- besitzt_Eigenschaft(kanarienvogel,Eigenschaft).
```

Wie das Prolog-Programm in Tabelle 1 illustriert, sind wissensbasierte Systeme im Unterschied zu herkömmlichen Programmen durch eine weitgehende Ex-

plizierung und Modularisierung ihrer Wissensbestände charakterisiert. Wichtig ist dabei, daß prozedurale Wissensbestände hinsichtlich ihrer Aktivierung a priori keinerlei Festlegungen unterworfen sind. Operatoren enthalten typischerweise die Prüfung von Bedingungen, unter denen sie prinzipiell anwendbar sind. Dadurch ergibt sich eine regelartige Struktur: ‚Wenn Bedingung, dann Aktion‘. Einzelne Wissenseinheiten werden aufgrund ihrer spezifischen Inhalte aktiviert, also aufgrund eines assoziativen Zugriffs und nicht entsprechend einer vorgegebenen starren Abarbeitungssequenz.

Eine zentrale Eigenschaft wissensbasierter Systeme im Sinne ablauffähiger Computerprogramme ist ihre **operative Interpretierbarkeit**. Es resultiert ein neuer Typ von Daten: Das beobachtbare Modellverhalten als Verhaltenssimulation, das ausschließlich auf der Grundlage des Modells abgeleitet wird. Mit anderen Worten formuliert: Wissensbasierte Systeme sind **Prozeß- und Performanzmodelle**, insofern sie die **Generierung von Modelldaten** ermöglichen.

Zusammenfassend läßt sich festhalten, daß wissensbasierte Systeme gekennzeichnet sind durch die Verwendung formaler syntaktischer Strukturen, die als repräsentations sprachliche Ausdrücke oder Symbolstrukturen interpretiert werden können. Sie repräsentieren als Propositionen, was das System über eine Anwendungsdomäne ‚weiß‘. Diese Sichtweise hat Smith (1982) in seiner Hypothese der **Wissensrepräsentation** (knowledge representation hypothesis) herausgestellt:

Any mechanically embodied intelligent process will be comprised of structural ingredients that a) we as external observers naturally take to represent a propositional account of the knowledge that the overall process exhibits, and b) independent of such external semantic attribution, play a formal but causal and essential role in engendering the behavior that manifests that knowledge.

(Smith, 1982, S.2)

### 3.2 Algorithmen, Berechenbarkeit und symbolverarbeitende Systeme

Aus der Sicht des Paradigmas der Symbolverarbeitung, das Newell als „the most fundamental contribution“ (1980, p. 136) ansieht, die von seiten der Informatik bzw. KI bisher in die Kognitionswissenschaft eingebracht wurde, können mentale Modelle mit Hilfe wissensbasierter Systeme rekonstruiert werden, wobei im Mittelpunkt der Betrachtung die Manipulation intern repräsentierter Symbolstrukturen steht.

Wo liegen die Ursachen für die Attraktivität dieser Betrachtungsweise? Zwei Gesichtspunkte sind wesentlich: Die Eigenschaften der **Immaterialität** und der **Berechnungsuniversalität** (vgl. auch Pylyshyn, 1989).

Analysieren wir zur Explikation dieser Eigenschaften einige Beispiele:

- Ist die Addition zweier natürlicher Zahlen ‚berechenbar‘?
- Ist die Klimaentwicklung ‚berechenbar‘?
- Sind menschliche kognitive Vorgänge wie Wahrnehmung, Problemlösen, Sprachverarbeitung ‚berechenbar‘?

Offenkundig hängt die Beantwortung dieser und vergleichbarer Fragen einerseits von unserem Verständnis des Begriffs ‚berechenbar‘ und andererseits der angesprochenen Vorgänge ab. Intuitiv unstrittig dürfte auf den ersten Blick allein die positive Beantwortung der ersten Frage sein. Die Addition zweier natürlicher Zahlen ist zweifellos berechenbar, aber warum bzw. in welcher Hinsicht eigentlich? Eine naheliegende, nichts desto trotz zu kurz greifende Antwort ist der Hinweis, daß es schließlich ‚Maschinen‘ und ‚Automaten‘ der unterschiedlichsten Art gibt, die eine Addition zweier natürlicher Zahlen durchführen können (z. B. Abakus, Taschenrechner, Computer). Damit Maschinen funktionieren, müssen sie zuvor entsprechend ‚instruiert‘ oder ‚programmiert‘ worden sein, d.h. entscheidend ist die Existenz (mindestens) eines Algorithmus, der genau angibt, wie eine Addition durchzuführen ist. Ein **Algorithmus** wird üblicherweise definiert als **ein durch eine endliche Vorschrift eindeutig festgelegtes effektives Verfahren zur Lösung einer Klasse von Problemen**. Ein Algorithmus akzeptiert beliebige zulässige Werte als Eingabe, verarbeitet diese wie in der Verfahrensbeschreibung festgelegt, so lange, bis das Ende der Berechnungen erreicht ist und das Ergebnis vorliegt. Effektivität meint dabei, daß ein Algorithmus - angewendet auf zulässige Eingabewerte - nach endlich vielen Schritten anhält und das ‚richtige‘ Ergebnis zurückliefert. Die eindeutige Festlegung durch eine endliche Vorschrift impliziert, daß ein Algorithmus in mechanischer Weise von einem symbolverarbeitenden System, einer Maschine oder einem Programm abgearbeitet werden kann.

Dabei ist ein **symbolverarbeitendes System** durch die folgenden grundlegenden Bestandteile definiert (Newell, 1980):

- eine **Menge primitiver Symbole** oder Zeichen (das ‚Alphabet‘ des Systems);
- eine **Menge von Regeln** zur Generierung komplexer, zusammengesetzter Symbolstrukturen oder Zeichenketten (die ‚Grammatik‘ des Systems);
- ein (oder mehrere) **Speicher**, in dem Symbole/Symbolstrukturen abgelegt und aus dem sie abgerufen werden können (das ‚Gedächtnis‘ des Systems);
- eine **Menge grundlegender Operationen** zur Verarbeitung der Symbole/Symbolstrukturen (etwa: Erzeugung, Zugriff, Veränderung, Löschung);
- ein (oder mehrere) **Prozessor(en)**, der die Arbeitsweise des Systems steuert und kontrolliert.

Nach dieser Definition sind symbolverarbeitende Systeme zunächst in einem abstrakten, theoretischen Sinn logisch-formale Systeme. Sie dürfen nicht mit real existierenden Rechenmaschinen, speziellen Programmiersprachen oder konkret implementierten Computerprogrammen verwechselt oder gleichgesetzt werden.

Die Konzipierung einer speziellen **Architektur des menschlichen kognitiven Apparats** als Spezialfall eines symbolverarbeitenden Systems ist eine zentrale Aufgabe der kognitionspsychologischen Theoriebildung. Eine **kognitive Architektur** ist ein spezifisches System von Speicherstrukturen („Gedächtnisse“) und Operationen („Verarbeitungsvorgänge“), das den Rahmen zur Rekonstruktion kognitiver Leistungen festlegt. Bekannte Beispiele kognitiver Architekturen sind das ACT\*-Modell von Anderson (1983) und das System SOAR (vgl. Laird, Rosenbloom & Newell, 1987; Newell, 1990). Eine allgemeine Diskussion des Architekturproblems, einschließlich einer vergleichenden Gegenüberstellung von ACT\* und SOAR findet sich bei Newell, Rosenbloom und Laird (1989).

Bisher wurde der Begriff der Berechenbarkeit informell verwendet. Ein Vorgang ist berechenbar, wenn es einen Algorithmus gibt, der von einem symbolverarbeitenden System ausgeführt werden kann. Um zu einer präzisen Operationalisierung des Algorithmusbegriffs zu gelangen, können drei Zugangsweisen unterschieden werden (vgl. Krämer, 1988, für eine geschichtlich orientierte Darstellung formaler Kalküle).

**Präzisierung mit Hilfe des mathematischen Funktionsbegriffs.** Die Arbeiten von Gödel (1931), Church (1936), Kleene (1936) und Turing (1936) auf dem Gebiet der mathematischen Logik legten die Grundlagen für diese Art der Präzisierung des Algorithmusbegriffs mit Hilfe des Konzepts der berechenbaren Funktionen. Sie führten zur Definition grundlegender Klassen berechenbarer Funktionen. Die wichtigsten Klassen sind die **elementaren Funktionen** (etwa: Identitäts-, konstante und Nachfolgefunktion), die **primitiv rekursiven Funktionen** und die **p-rekursiven Funktionen**. Umfassende Darstellungen der mathematischen Berechenbarkeits- oder Rekursionstheorie finden sich etwa in den Lehrbüchern von Minsky (1967) oder Rogers (1987).

**Präzisierung mit Hilfe formaler Sprachen/Grammatiken.** Bei dieser Betrachtungsweise, die auf Autoren wie Post (1943), Markov (1954) oder Chomsky (etwa Chomsky, 1963) zurückgeht, werden die Ein- und die Ausgabe von Algorithmen als Worte eines jeweils passenden Alphabets aufgefaßt. Ein Algorithmus ist dabei charakterisiert als eine Menge von Regeln zur schrittweisen Manipulation von Zeichenreihen. Man unterscheidet zwischen **regulären, kontextfreien, kontextsensitiven** und **allgemeinen Sprachen/Grammatiken**.

**Präzisierung mit Hilfe von Klassen von Maschinen.** Die Informatik kennt und studiert eine Reihe wichtiger Klassen von Maschinen oder Automaten als Spezialfälle (abstrakter) symbolverarbeitender Systeme (vgl. Hopcroft & Ullman, 1979; Stetter, 1988). Insbesondere sind dies die Klassen der **endlichen Automaten**, der **Kellerautomaten**, der **linear beschränkte Automaten** und - als die wohl bekannteste Klasse - die **Turingmaschinen**.

Im Grunde handelt es sich bei der Turingmaschine um ein **abstraktes Gedankenmodell**, das die Durchführung theoretischer Analysen zur Beantwortung grundlegender Fragen über die Natur von Rechenprozessen erlaubt. Für praktische Berechnungen ist sie nicht geeignet. Die Bedeutung der Turingmaschine für die Theorie der Berechenbarkeit ergibt sich daraus, daß sie - einen genügend großen, aber endlichen Zeitraum vorausgesetzt - jeden Algorithmus und damit jede Berechnung durchführen kann. Diese **Eigenschaft der Berechnungsuniversalität** und die damit einhergehende Definierbarkeit der algorithmischen Berechenbarkeit, unabhängig von speziellen Klassen von Maschinen, ist von fundamentaler Bedeutung. Einerseits beruht darauf die umfassende Anwendbarkeit programmierbarer Rechenmaschinen zur algorithmischen Lösung der unterschiedlichsten Probleme; andererseits gestattet sie, den Nachweis der algorithmischen Lösbarkeit eines Problems anhand der am besten geeigneten Klasse von Maschinen zu führen.

Anzumerken bleibt, daß von den oben genannten Klassen von Maschinen lediglich die Turingmaschine die Eigenschaft der Berechnungsuniversalität besitzt. Die übrigen Klassen von Maschinen sind weniger leistungsfähig. Für den Zusammenhang von Sprachen/Grammatiken und Maschinentypen gilt: Reguläre Sprachen sind äquivalent zu endlichen Automaten, kontextfreie bzw. kontextsensitive Sprachen sind äquivalent zu Kellerautomaten bzw. zu linear beschränkten Automaten und allgemeine Sprachen sind äquivalent zu Turingmaschinen.

Allerdings bedeutet Berechnungsuniversalität keineswegs, daß alle Vorgänge berechenbar sind. Es gibt Probleme, die **nicht** algorithmisch gelöst werden können. Bekanntestes Beispiel für diese Form der Unvollständigkeit formaler Systeme ist das sog. Halteproblem. Man kann zeigen, daß es keinen Algorithmus gibt, der ‚berechnet‘, ob eine beliebige Turingmaschine anhält, also nach endlich vielen Schritten stehenbleibt oder nicht (vgl. Rogers, 1987, weitere Beispiele nicht berechenbarer Funktionen findet man in Brauer, 1990).

Bei vielen Fragestellungen kommt es für die Frage der Berechenbarkeit darüber hinaus entscheidend auf den **Auflösungsgrad** an, mit dem ein Vorgang betrachtet wird. Berechenbarkeit auf einem geringen Auflösungsgrad impliziert dabei keineswegs Berechenbarkeit auf einem höheren Auflösungsgrad. So bedeutet etwa eine systematische Regelmäßigkeit in aggregierten Daten nicht unbedingt eine Vorhersagbarkeit des Einzelfalls. Je ‚genauer‘ die Vor-

hersage der künftigen Entwicklung sein soll, desto eher wird man den zugrundeliegenden Vorgang als nicht berechenbar, bzw. als beim gegenwärtigen Kenntnisstand nicht berechenbar bezeichnen. Beispielsweise gilt hinsichtlich der Frage nach der Berechenbarkeit langfristiger Klimaentwicklungen: Modelle einer globalen Klimaveränderung existieren, aber sie erlauben keine präzisen Aussagen über lokale Veränderungen. Dabei geht die Meteorologie durchaus von der Annahme aus, daß die Entwicklung des Klimas global und lokal systematischen Regelmäßigkeiten gehorcht, die man zwar im Detail noch nicht kennen mag, aber prinzipiell ‚entdecken‘ und mit Hilfe geeigneter Modelle rekonstruieren könnte. Berechenbarkeit und Vorhersagbarkeit können in diesen und vergleichbaren Fragestellungen auch als Relation zwischen dem vorherzusagenden Geschehen und dem aktuellen Wissen über eben dieses Geschehen aufgefaßt werden. Danach wird ein Geschehen als ‚nicht berechenbar‘ bezeichnet, wenn kein ausreichendes Wissen zu seiner Vorhersagbarkeit vorhanden ist.

Wie sieht es aber nun mit der Berechenbarkeit kognitiver Leistungen des Menschen aus? Zur Beantwortung dieser Frage wird mit Bezug auf Newell (1982) eine Mehrebenenbetrachtung von Theorien der Informationsverarbeitung eingeführt.

### 3.3 Mehrebenenbetrachtung von Theorien der Informationsverarbeitung

Das Verhalten, die Funktionalität und die Struktur vieler technischer, aber auch biologischer Systeme lassen sich auf verschiedenen Ebenen beschreiben, analysieren und erklären. Häufig liegt das Problem gerade ‚darin‘, die richtige Ebene der Beschreibung und Erklärung zu finden. Welche Ebenen sind zu unterscheiden, wenn menschliche kognitive Aktivitäten mit Hilfe wissensverarbeitender Systeme rekonstruiert werden?

Dazu wurden in den letzten Jahren verschiedene Vorschläge gemacht, insbesondere von Marr (1982), Newell (1982) und Pylyshyn (1985), aber auch von Anderson (1990). Wegen ihres nachhaltigen Einflusses auf die gesamte Grundlagendiskussion sollen die Überlegungen von Newell im folgenden kurz vorgestellt werden.

**Newells Mehrebenenbetrachtung.** Ausgangspunkt für Newells Überlegungen war die Frage, welcher ‚Wissensbegriff‘ der KI überhaupt zugrundeliegt und damit insbesondere das Problem, welcher Zusammenhang zwischen Wissen und seiner Rekonstruktion in formalen Strukturen besteht (Newell, 1982; vgl. auch Daiser, 1984). Newell unterscheidet drei grundlegende Ebenen der Betrachtung, die **Wissensebene**, eine **Programm- bzw. Symbolebene** und eine phy-

**sikalische Ebene.** Die einzelnen Ebenen bilden dabei für Newell eine Hierarchie mit zwei wesentlichen Eigenschaften. Sie sind autonom definierbar, aber dennoch gibt es Abhängigkeitsbeziehungen zwischen ihnen in dem Sinne, daß jede Ebene auf der unmittelbar unter ihr angesiedelten Ebene aufsetzt. Dies impliziert, daß jede Ebene - zumindestens im Prinzip - auf die unter ihr liegende Ebene zurückgeführt bzw. ‚reduziert‘ werden kann. Eine derartige, **analytisch-reduktionistische Vorgehensweise** ist eine Voraussetzung für die Bildung von Simulationsmodellen. Wichtig ist in diesem Zusammenhang, daß diese Reduktion keineswegs eindeutig ist. So kann etwa eine theoretische Vorstellung in der Regel auf unterschiedliche Weise formalisiert werden. Ebenso kann ein Algorithmus in verschiedener Weise implementiert werden.

Erklärungen der Funktionalität und des Verhaltens eines Systems sind jeweils innerhalb einer Ebene zu entwickeln. Eine Vermischung von Betrachtungsebenen impliziert keine besseren Erklärungschancen, sondern führt zu begrifflicher Verwirrung! Zu Recht wird beispielsweise immer wieder darauf hingewiesen, daß theoretische Vorstellungen einerseits und Computerprogramme andererseits nicht miteinander verwechselt werden dürfen (vgl. Herrmann, 1982). Darüber hinaus stehen Vorstellungen, die sich auf verschiedene Ebenen eines Systems beziehen, nicht unmittelbar in Konkurrenz zueinander - im Unterschied zu solchen Vorstellungen, die beanspruchen, ein System innerhalb ein- und derselben Ebene zu erklären.

Kehren wir zu den Überlegungen Newells zurück. Oberhalb der Programm- oder Symbolebene liegt als ‚oberste‘ Ebene die Wissensebene. Ihre Existenz wird von Newell als Hypothese formuliert (**knowledge level hypothesis**). In Abhebung von den übrigen Ebenen lassen sich die Besonderheiten der Wissensebene in zwei charakteristischen Eigenschaften zusammenfassen. Erstens ist die Wissensebene nach Newell auf eine radikale Art unvollständig. Dies impliziert, daß Verhalten nicht vollständig beschreibbar und vorhersagbar ist.

Sometimes behavior can be predicted by the knowledge level description; often it cannot. The incompleteness is not just a failure in certain special situations or in some small departures. The term radical is used to indicate that entire ranges of behavior may not be describable at the knowledge level, but only in terms of systems at a lower level (namely, the symbolic level). However, the necessity of accepting this incompleteness is an essential aspect of this level.

(Newell, 1982, S. 105)

Zweitens ist Wissen nach Newell ein Medium, das keinerlei spezifische materielle Zustandsbeschreibung besitzt, sondern lediglich funktional definiert werden kann.

**Knowledge.** Whatever can be ascribed to an agent, such that its behavior can be computed according to principle of rationality.

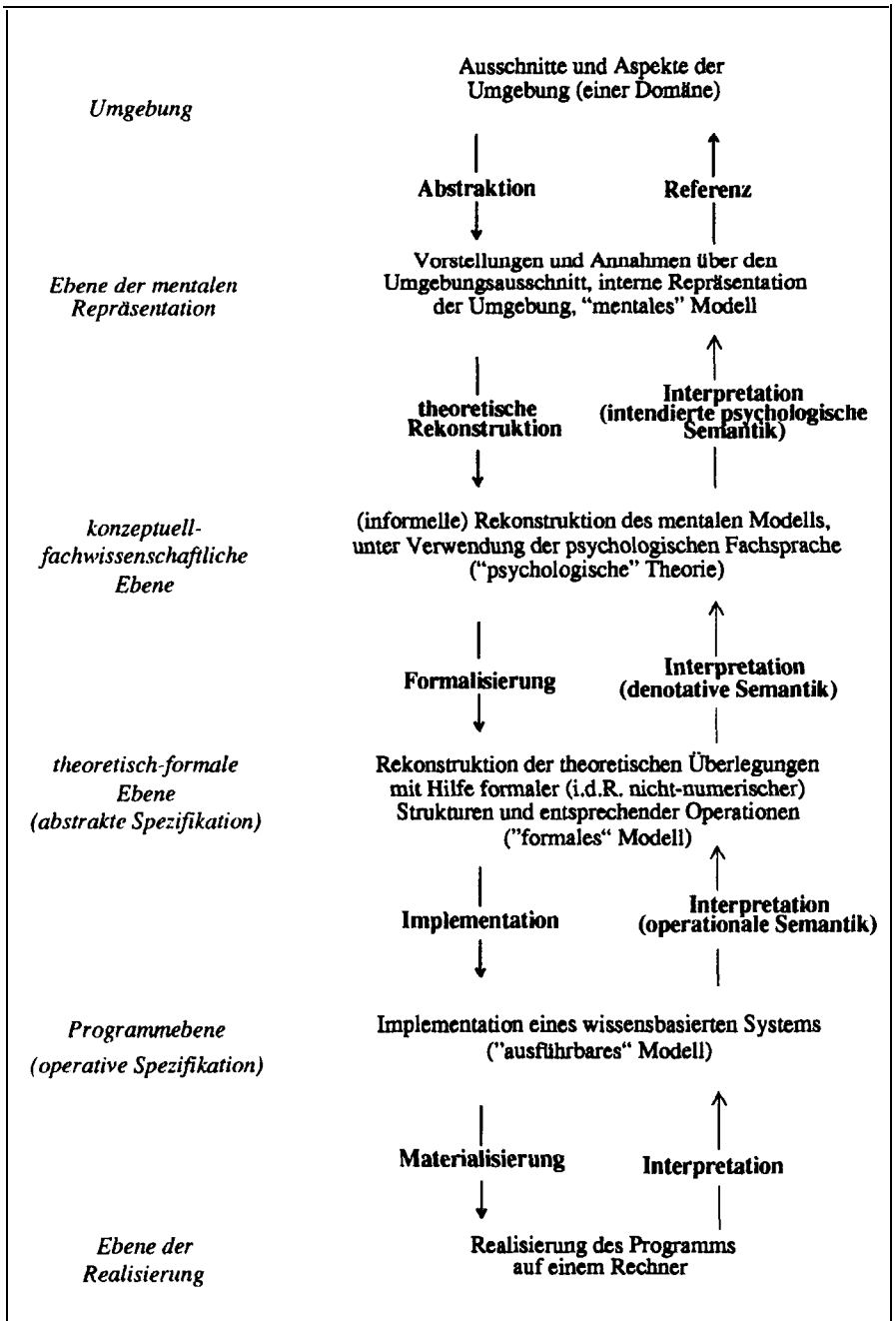


Abb. 2: Mehrebenenbetrachtung von Theorien der Informationsverarbeitung aus psychologischer Sicht

Knowledge is to be characterized entirely **functionally**, in terms of what it does, not **structurally**, in terms of physical objects with particular properties and relations.

(Newell, 1982, S.105)

Wissen wird auf diese Weise zu einem Kompetenzbegriff im Sinne eines Potentials, das verhaltenswirksam werden kann, aber nicht muß und das nicht passiv, sondern nur aktiv im Prozeß der Auswahl und Anwendung realisiert ist. Eine indirekte und approximate materielle Struktur bekommt Wissen erst durch seine Rekonstruktion in Form wissensverarbeitender Systeme auf der Programm- oder Symbolebene.

Abbildung 2 faßt einige Überlegungen aus diesem und den vorangehenden Abschnitten in Form einer Mehrebenenbetrachtung von psychologischen Theorien der Informationsverarbeitung zusammen.

Die **Ebene der mentalen Repräsentation** stellt eine Rekonstruktion von Ausschnitten einer tatsächlichen oder einer vorgestellten **Umgebung** dar. Sie ist eine **Abstraktion**. Umgekehrt bezeichnen die mit dem mentalen Modell angesprochenen Vorstellungen Objekte der Erfahrung. Diese Vorstellungen erlauben das Wiedererkennen von und das gedankliche Operieren mit Objekten der Erfahrung. Diese Beziehung wird **Referenz** genannt.

Die **theoretische Rekonstruktion** des mentalen Modells einer Person unter Verwendung informeller, theoretischer psychologischer Begriffe charakterisiert die **konzeptuell-fachwissenschaftliche Ebene**. Die Interpretation des damit angesprochenen psychologischen Modells erster Art entspricht der **intendierten psychologischen Semantik**.

Die unteren drei Ebenen in Abbildung 2 ergeben sich aus der Verwendung wissensbasierter Systeme zur Rekonstruktion mentaler Modelle. Eine **Formalisierung** der in der Regel natürlich-sprachlichen psychologischen Begriffe führt zur **theoretisch-formalen Ebene**. Ihre **Implementation** entspricht einer Transformation in eine operative Spezifikation auf der **Programmebene**, deren **Materialisierung** schließlich zur **physikalischen Realisierung** des Programms auf einem speziellen Rechner führt. Umgekehrt betrachtet kann eine operative Spezifikation als eine **Interpretation** ihrer jeweiligen Materialisierung aufgefaßt werden. Die Interpretation eines Programms im Sinne seiner **operationalen Semantik** führt zu einer abstrakten Spezifikation, deren **denotative Semantik** im vorliegenden Kontext mit einer Abbildung formalisierter Strukturen und Operationen auf eine psychologische Begrifflichkeit korrespondiert.

Diese Mehrebenenbetrachtung von Theorien der Informationsverarbeitung ist eine idealisierte Betrachtung. Dies wird sichtbar, wenn man die in Abbildung 2 eingeführten Ebenen mit der von Newell entwickelten Betrachtung in Beziehung setzt. Insbesondere die Zuordnung der Wissensebene bereitet Schwierigkeiten. Ihre Funktionalität wird sowohl durch die Ebene der mentalen Re-

Präsentation als auch durch die konzeptuell-fachwissenschaftliche Ebene abgedeckt. Eine saubere Trennung zwischen einer kognitionspsychologischen Begrifflichkeit, die durch sprachliche Ausdrücke wie Ziele, Absichten, Wissen, Pläne und dergleichen gekennzeichnet ist und implementationsorientierten Begriffen wie deklarative und prozedurale Wissensbestände, Kontrollstrategien usw. ist häufig nicht einfach, wenn auch grundsätzlich wünschenswert.

#### **4. Bewertung wissensbasierter Systeme als psychologische Modelle**

Computer science has given psychology a new way of expressing models of cognition that is much more detailed and precise than its predecessors. But unfortunately, the increased detail and precision in stating models has not been accompanied by correspondingly detailed and precise arguments **analyzing and supporting** them.

(VanLehn, Brown & Greeno, 1984, S.237)

In diesem Abschnitt untersuchen wir wissensbasierte Systeme zur Modellierung kognitiver Strukturen und Prozesse anhand der Bewertungskriterien, die üblicherweise an Theorien herangetragen werden. Dazu zählen etwa (vgl. Gadenne, Theoriebewertung, in diesem Band): Die logische Konsistenz und Widerspruchsfreiheit der Theorie, die Einheitlichkeit des Begriffssystems (semantische Einheitlichkeit), der Informationsgehalt und die Tiefe der Theorie, ihre praktische Relevanz und - besonders herausgehoben - ihre empirische Adäquatheit.

##### 4.1 Empirische Adäquatheit

Die empirische Adäquatheit einer Modellierung ist das zentrale Gütemerkmal zur Beurteilung einer psychologischen Theorie. Damit ist auch ein Kriterium angesprochen, das wie kein anderes zwischen einer psychologisch motivierten kognitionswissenschaftlichen Forschung, bei der der Mensch das Objekt der Theoriebildung ist, und dem Forschungsansatz der künstlichen Intelligenz als Teil der Informatik unterscheidet. Aus computerwissenschaftlicher Sicht steht der Entwurf und die Implementation eines intelligenten Systems mit bekannten formalen Eigenschaften und definierten Leistungen im Vordergrund. Die Frage einer psychologisch validen Rekonstruktion menschlicher intelligenter Leistungen ist dabei nicht von primärem Interesse.

##### **4.1.1 Eine theoretische Rahmenvorstellung**

Zur Behandlung der Frage der empirischen Adäquatheit einer kognitiven Modellierung greifen wir auf die theoretische Rahmenvorstellung zurück, die

schon im zweiten Abschnitt skizziert wurde. Sie ist in vereinfachter und zugleich für die hier interessierenden Fragen zugeschnittener Form in Abbildung 3 dargestellt.

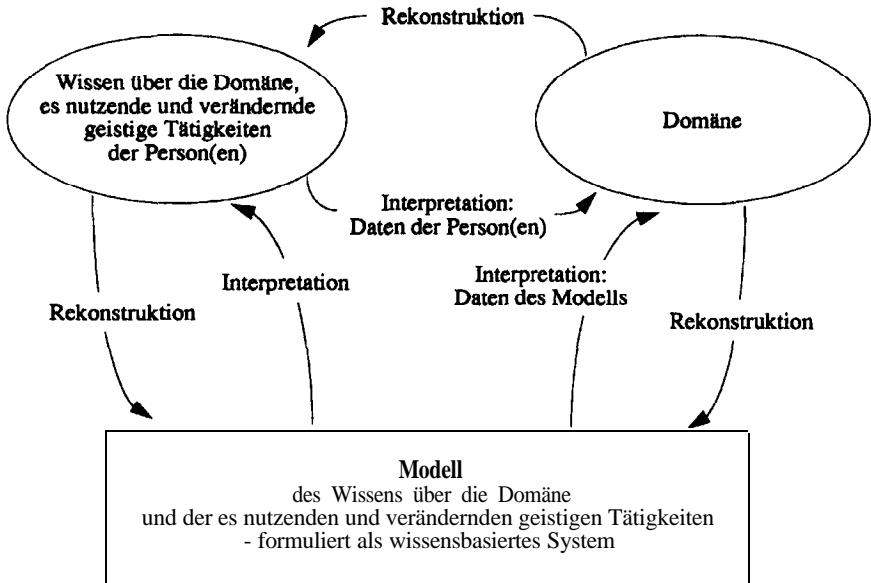


Abb. 3: Bewertung wissensbasierter Systeme: Vergleich von Personen- und Modell-daten zur Überprüfung der empirischen Adäquatheit

Eine Person (oder eine Gruppe von Personen) beschäftigt sich mit einer bestimmten Domäne, einem Ausschnitt der Umgebung. Das Ergebnis der Rekonstruktion ist Wissen über die Domäne. Über die es nutzenden und verändernden geistigen Tätigkeiten resultieren in Interaktion mit der Domäne die Daten der Person(en). Die psychologisch valide Rekonstruktion dieses Wissens und der auf ihm ablaufenden Prozesse ist Ziel der Modellierung, hier seiner Formulierung als wissensbasiertes System. Dieses System soll die Domäne in derselben Weise rekonstruieren, wie dies die betrachteten Personen tun. Überprüfbar ist dies anhand der Interpretation des Modells, d.h. über die Daten, die das Modell in Interaktion mit der Domäne liefert. Modell- und Personendaten stellen die Basis für die Prüfung der empirische Adäquatheit der Modellierung dar.

Wir haben bisher nicht spezifiziert, ob sich die Modellierung nur auf eine Person oder alle Personen einer bestimmten Population beziehen soll. Tatsächlich ist dies ein Punkt, der in der Literatur nicht selten für Verwirrung sorgt. Es ist zu unterscheiden, ob ein idiographisches Modell, das eine bestimmte Person beschreibt, ein für alle Personen einer Population prototypisches Modell oder ein individualisiertes Modell vorliegt, das von allgemeinen, auf alle Personen

**einer Population zutreffenden Annahmen** ausgeht, **darüber hinaus aber inter-individuell unterschiedliche Annahmen und Komponenten zuläßt.**

Wissenschaftstheoretisch gesehen sind Modellierungen des letzten Typs am vorteilhaftesten. In ihnen wird klar expliziert, was als allgemein und was als individuell spezifisch angenommen wird. Interindividuelle Varianz wird theoretisch begründet. Individualisierte Modelle erfordern eine Diagnosekomponente, um die individualspezifischen Anteile festlegen zu können. Anhand einer Einzelfallanalyse entwickelte Modellierungen beschreiben zumeist die mentalen Strukturen und Prozesse nur des einen betrachteten Individuums. Allerdings weisen sie oft keine klare Trennung von allgemeinen und personenspezifischen Annahmen auf. Sie stellen unter dem Aspekt der Verallgemeinerbarkeit einen sehr problematischen Fall dar.

In den Tabellen 2 und 3 werden mit den Systemen MULEDS (vgl. Plötzner, Spada, Stumpf & Opwis, 1990) und KAGE (vgl. Plötzner, 1990; Plötzner & Spada, 1992) zwei Beispiele von Modellierungen skizziert, von denen in der einen interindividuelle Unterschiede zugelassen sind, die andere unter Ausklammerung interindividueller Varianz Wissensveränderungen, also ein prototypisches, hier aber ‚dynamisches‘ mentales Modell rekonstruiert. Wir werden in den nachfolgenden Ausführungen zu Fragen der empirischen Prüfung verschiedentlich Bezug auf diese Modellierungen nehmen.<sup>2</sup>

Tabelle 2: Programmsystem MULEDS (Multi-Level-Diagnosis-System)

---

Theoretische Grundlage des mit MULEDS realisierten Diagnosemodells ist die Rahmenvorstellung einer multiplen mentalen Repräsentation von Wissen über physikalische Gegenstandsbereiche entlang der Dimension qualitativ/quantitativ. Methodisch ist MULEDS dem im Zusammenhang mit der Diagnose und Modellierung arithmetischer Fertigkeiten bekannt gewordenen ‚Fehlerbibliotheksansatz‘ zuzuordnen. Zentral sind hierbei zwei Mengen von im Modell vorselektierten Wissenselementen. Erstens, die Menge korrekter Wissenselemente, die das zu einer erfolgreichen Problemlösung benötigte bereichsspezifische Wissen repräsentieren. Zweitens, eine Menge fehlerhafter bzw. unvollständiger Wissenselemente, die typische Denkfehler und/oder Mißkonzepte widerspiegeln.

Es wurden korrekte, fehlerhafte und unvollständige Wissenskomponenten auf jeweils einer qualitativ-relationalen, einer quantitativ-relationalen und einer quantitativen Repräsentationsebene zu den Phänomentypen des zentralen und schiefen Stoßes, eines Teilgebiets der klassischen Mechanik formuliert. Zusätzlich wurde mathematisches/experimentelles Wissen, das der Verknüpfung mehrerer Vorhersagen dient, in Form korrekter Regeln repräsentiert. Zur Implementation wurde der Produktionssysteminterpretierer PRISM verwendet.

Bei Anwendung von MULEDS auf einen Datensatz von Personen, die entsprechende Aufgaben bearbeitet haben, resultieren im positiven Fall ausführbare individuelle Wissensmodelle. Im negativen Fall ist keine Diagnose des Wissens der Person möglich.

---

<sup>2</sup> Diese Arbeiten erfolgten im Rahmen von Forschungsvorhaben, die durch die Deutsche Forschungsgemeinschaft unterstützt wurden (Sp 251/2-x; Sp 251/6).

Tabelle 3: Programmsystem KAGE (Knowledge Acquisition Governed by Experimentation)

---

Basierend auf der Theorie multipler mentaler Repräsentationsebenen (Plötzner et al., 1991) und unter Verwendung analysebasierter Lernmechanismen wird der Erwerb von Wissen über funktionale Beziehungen rekonstruiert. Modelliert wird der Lernprozeß von Schülern, die im Rahmen einer Mikrowelt-Lernumgebung simulierte physikalische Experimente durchführen. Zentrale Annahme ist, daß ein Schüler zunächst mit Hilfe experimentellen und heuristischen Wissens versucht, anhand einzelner Experimente Hypothesen darüber zu formulieren, welche physikalischen Größen wie zusammenhängen. Die anschließende mathematische Elaboration der Art des Zusammenhangs wird dann im Sinne eines Analogieschlusses auf neue experimentelle Situationen übertragen. Solche Analogieschlüsse sind im Modell sowohl innerhalb als auch zwischen Repräsentationsebenen vorgesehen.

Zu den rekonstruierten kognitiven Mechanismen zählen: Identifikation funktional abhängiger Variablen; analoges Schließen unter Heranziehung analoger Fälle, Elaboration des betrachteten Falls, Extrapolation auf den neuen und Evaluation der Inferenz, Abstraktionsbildung und Ableitung neuer experimenteller Vorhersagen. KAGE behandelt insbesondere proportionale und umgekehrt proportionale Beziehungen, und zwar auf verschiedenen Repräsentationsstufen, die sich auf einer Dimension qualitativ/quantitativ anordnen lassen. Im Vordergrund der Modellierung steht die jeweilige Nutzung von Wissen, das auf einer informationsärmeren Repräsentationsebene erworben wurde, zur Formulierung von Hypothesen auf der nächst höheren Ebene.

KAGE liegt implementiert in PROLOG vor.

---

#### **4.1.2 Probleme**

Auf den ersten Blick erscheint die empirische Geltungsprüfung von als wissenschaftsbasierte Systeme realisierten Modellierungen problemlos. Offensichtlich gilt es, die Daten der Personen, also die empirischen Daten, mit den vom Modell generierten Daten zu vergleichen. So könnte man etwa die mit Hilfe des Systems MULEDS verschiedenen Schülern zugeschriebenen Wissensmodelle zur Generierung von Antworten auf weitere Aufgaben heranziehen, um dann diese Modelldaten mit den von den Schülern selbst gezeigten Reaktionen zu vergleichen. Eine Prüfung des Systems KAGE könnte wie folgt ablaufen: es wird eine bestimmte Aufgabensequenz sowohl Schülern wie auch dem System vorgelegt. Die aus der Bearbeitung der Aufgaben resultierende Veränderung des Wissens der Schüler bzw. ihre damit korrespondierenden Antworten werden mit den durch das Modell vorhergesagten Antworten in Beziehung gesetzt. Ähnlich könnte bei anderen Modellierungen, bei denen mit wissenschaftsbasierten Systemen gearbeitet wird, vorgegangen werden. Geltungskriterium ist die Passung von Personen- und Modelldaten.

Dieses Vorgehen ist möglich, da die Arbeit mit wissenschaftsbasierten Systemen zu einem neuen Datentyp, den anhand der Modellierung selbst generierten Daten, führt. Aus der Modellierung resultiert eine bestimmte Form von Perfor-

manz. Sie ergibt sich aus einem besonderen Merkmal dieser Forschungsmethodik, der ‚Lauffähigkeit‘ wissensbasierter Systeme. Ein Computermodell liefert nicht nur eine Beschreibung von Aktivitäten, sondern setzt im Sinne einer Kompetenzmodellierung die Spezifizierung generativer Mechanismen voraus, auf deren Grundlage die in Frage stehenden Verhaltensphänomene generiert werden können. Dies macht ja gerade einen wichtigen Aspekt der Faszination dieses Forschungsansatzes aus. Der Wissenschaftler wird in die Lage versetzt, (Modell-) Verhalten zu generieren und zu manipulieren. Wo liegen aber nun die Probleme?

Ein zentrales theoretische Problem ist die **Unterdeterminiertheit einer Modellierung durch empirische Daten**. Nicht minder gravierend ist das auf den ersten Blick gegensätzliche Problem, daß **Modelldaten und empirische Daten** bei strenger Prüfung fast immer **differieren**.

Was an diesen Problemen ist aber nun der Modellierung in Form von wissensbasierten Systemen anzulasten, und was gilt in gleichem bzw. in noch viel höherem Maße auch für andere Formen der Theorieformulierung?

Modelle - welcher Art auch immer - sind prinzipiell empirisch unterdeterminiert. Eine perfekte Passung von Modell und Daten ist nie zu erwarten. Bei Modellierungen des hier betrachteten Typs treten diese Probleme aber aus verschiedenen Gründen besonders deutlich hervor.

(1) Es werden nicht nur Oberflächenmerkmale geistiger Tätigkeiten (wie etwa Reaktionszeiten) beschrieben, sondern es werden die Strukturen und Prozesse modelliert, die den beobachteten Phänomenen als zugrundeliegend angenommen werden. VanLehn et al. (1984) unterscheiden im Anschluß an Moravcsik (1980) in diesem Zusammenhang zwischen ‚**tiefen**‘ und ‚**oberflächlichen**‘ Theorien. Oberflächliche Theorien nehmen ausschließlich Bezug auf die beobachtbaren Ergebnisse psychischer Vorgänge und setzen diese zumeist korrelativ miteinander in Verbindung. Im Unterschied dazu versuchen tiefe Theorien die ‚black box‘ zwischen Eingangs- und Ausgangsgrößen zu öffnen und jene mentalen Strukturen und Prozesse zu beschreiben, auf die die beobachtbaren Phänomene zurückgeführt werden können.

(2) Computersimulierte Theorien weisen einen sehr hohen Informationsgehalt auf, da sie viele und sehr präzise formulierte Aussagen enthalten. Die Kehrseite tiefer Theorien ist ein ungünstiges Verhältnis zwischen Postulaten und empirischen Daten.

(3) Hinzu kommt, daß zur Gewährleistung der Lauffähigkeit eines wissensbasierten Systems Annahmen in einem Detaillierungsgrad getroffen werden müssen, der den wissenschaftlichen Kenntnisstand zu dem entsprechenden Phänomenbereich häufig weit übersteigt. Im Vergleich zu Theorien, die nur den Effekt bestimmter Bedingungen beschreiben und vorhersagen, erfordert die Modellierung mentaler Prozesse wesentlich differenziertere Annahmen.

(4) Darüber hinaus wird mit Hilfe wissensbasierter Systeme meistens Einzelverhalten generiert. Der Vergleich von Modell- und Personendaten erfolgt dann häufig auch auf dieser Ebene. Es wird nicht mit über alle Personen gemittelten Lösungszeiten, mit relativen Häufigkeiten richtiger Antworten oder ähnlichen aggregierten Daten und damit einem reduzierten Auflösungsgrad gearbeitet. Diese Aussage gilt übrigens auch für Modellierungen, die mit dem Ziel entwickelt werden, prototypisch für alle Personen einer Population zu sein. Hingegen ist die Verwendung aggregierter Daten die Regel bei der statistischen Prüfung verbal formulierter Theorien, aber auch mathematisch-numerischer Modelle. So schreiben Erdfelder und Bredenkamp (in diesem Band):

Empirische Untersuchungen in der Psychologie zeichnen sich fast ausnahmslos dadurch aus, daß substanzwissenschaftliche Fragestellungen mittels statistischer Hypothesentests beantwortet werden. Bemerkenswert ist dieses Vorgehen vor allem deshalb, weil die interessierende psychologische Hypothese und die letztlich getestete statistische Hypothese im Regelfall nicht identisch sind. Psychologische Hypothesen beziehen sich typischerweise auf psychische Vorgänge, die im einzelnen Individuum ablaufen... Statistisch getestet werden aber . . . häufig Populationsaussagen . . .

(5) Ein weiterer Aspekt ist, daß wissensbasierte Systeme zumeist deterministisch formuliert sind. Bei Theorien, die primär Oberflächenmerkmale beschreiben, kann man eine zusätzliche Flexibilität - man spricht auch von einer Robustheit der Theorie gegenüber Falsifikationsbemühungen - dadurch erzielen, daß nur Wahrscheinlichkeitsaussagen formuliert werden. Die Explizitheit der Formulierung wissensbasierter Systeme in Verbindung mit dem Ziel der Generierung von Verhalten läßt probabilistischen Formulierungen wenig Raum. Eine hohe Falsifikationswahrscheinlichkeit ist die negative (oder positive?) Folge.

Tiefe, deterministisch formulierte Theorien in Form wissensbasierter Systeme, die Einzelverhalten generieren, sind empirisch leicht angreifbar. Sind sie damit aber weniger adäquat als die auf einem sehr eingeschränkten Auflösungsgrad formulierten oberflächlichen Theorien?

Um diese Frage beantworten zu können, ist die Prüfung der empirischen Adäquatheit in einen größeren Zusammenhang zu stellen. Wir werden die Rolle von Empirie im Prozeß der Theorieentwicklung aus drei Blickwinkeln betrachten: Empirie als Vorbild für die Modellierung, empirische Geltungsprüfung anhand einer Analyse der Modelldaten selbst, und anhand eines methodisch reflektierten Vergleichs von Modell- und Personendaten.

### **4.1.3 Empirie als Vorbild**

Stellen wir zunächst die Rolle der Empirie als Richterin über die Adäquatheit einer Modellierung zurück. Betrachten wir empirisches **Wissen als Vorausset-**

**zung und Leitfaden für die Modellkonstruktion.** Diese Phase der Theorieentwicklung findet häufig kaum Beachtung in methodischen Erörterungen im Fach Psychologie (vgl. aber Dörner, 1989; in diesem Band).

The function of empirical observations in this research programme is different from its role in traditional experimental psychology. In a traditional study, the purpose of observing a person's performance is to ascertain the effect of some stimulus variable on his responses. In the simulation programme, the function of empirical investigations is, broadly speaking, to describe human Performance in as much detail as possible, in order to constrain model building. . . . knowledge takes the form of **know-how**.

(Ohlsson, 1988, S. 14)

Nach Ohlsson kommt somit empirischen Erhebungen für die Theorieformulierung zur Ableitung von Designentscheidungen bei der Entwicklung wissensbasierter Systeme eine zentrale Rolle zu. Jeder kognitiven Modellierung muß eine intensive empirische Analyse der „Sachverhältnisse“ (Selz, 1913) des betrachteten Phänomenbereichs vorausgehen. Dabei handelt es sich um ein eher interpretatives Vorgehen (vgl. Soeffner & Hitzler, in diesem Band) mit einem geringen Grad an Datenrestriktion. Selbstverständlich werden auch verbal formulierte Theorien und mathematisch-numerische Modelle unter Heranziehung des Vorwissens über die Domäne gestaltet. Merkmale, wie Exaktheit, Detailliertheit und Lauffähigkeit eines wissensbasierten Systems, zwingen aber zu einer wesentlich intensiveren Auseinandersetzung mit dem Gegenstandsbereich und damit zu einer umfassenden Datenerhebung im Vorfeld der Modellierung.

Jeder Schritt in der Modellierung, jede getroffene Annahme ist theoretisch und empirisch zu motivieren und zu begründen. Auf diese Weise wird ein hoher empirischer Gehalt der Modellierung schon in der Phase der Entwicklung des wissensbasierten Systems sichergestellt. Als Beispiel können die Arbeit von Chi, Bassok, Lewis, Reimann und Glaser (1989) und die auf ihr aufbauenden Modellierungen herangezogen werden. Chi et al. (1989) zeigen, daß Schüler insbesondere dann aus vorgegebenen Beispielen lernen, wenn sie die Beispiele reflektierend nachvollziehen und sich selbst erklären. Ansätze einer Modellierung dieses Selbsterklärungseffekts werden von Reimann und Schult (1991) sowie VanLehn, Jones und Chi (1992) diskutiert.

#### **4.1.4 Modellperformanz als Datenquelle**

Ein wichtiger Teil der Geltungsprüfung besteht in einer systematischen Analyse der Modelldaten vor dem Hintergrund des Wissens über den betrachteten Phänomenbereich. Folgende Fragen sind zu prüfen:

- Rekonstruiert die Modellierung die zu erklärenden Phänomene des untersuchten Gegenstandsbereichs (situative Variation) und für den analysierten Personenkreis (individuelle Variation)?
- Generiert die Modellierung auch unsinnige, bisher nicht empirisch beobachtete Phänomene, sog. Sternchen-Daten im Sinne von VanLehn et al., 1984?
- Führen gezielte Eingriffe in das Modell zu den theoretisch erwarteten Änderungen des Modellverhaltens?

Die erste Frage bezieht sich auf die **Suffizienz** einer Modellierung. Es wird geprüft, ob das Modell als hinreichend zur Erzeugung des interessierenden Verhaltens angesehen werden kann. Dabei ist insbesondere zu untersuchen, ob die Phänomene des Gegenstandsbereichs in ihrer gesamten Breite generiert werden können und ob die Modellierung der in der analysierten Population beobachteten interindividuellen Varianz gerecht wird. Diese Prüfung beschränkt sich nicht nur auf die Rekonstruktion korrekten Verhaltens, sondern umfaßt auch fehlerhafte Reaktionen.

Die Prüfung der Suffizienz ist ein strenger Falsifikationstest. Ist ihr Ergebnis negativ, kann auf weitere Schritte der Geltungskontrolle verzichtet werden. Umgekehrt bedeutet ein positives Ergebnis nicht notwendigerweise, daß die Modellierung auch empirisch adäquat ist. In diesem Fall ist ‚nur‘ gesichert, daß die Modellierung hinreichend zur Verhaltensgenerierung ist.

Die Analyse der Frage, ob ein wissenschaftliches System auch empirisch **nicht** beobachtbares Verhalten zeigt, stellt eine weitere Möglichkeit der Falsifizierung dar. Sie wurde von VanLehn (1983, Brown & VanLehn, 1980) im Zusammenhang mit der Entwicklung und Testung ihrer Flickwerktheorie (repair-theory) in die Diskussion eingeführt. Geprüft wurde, ob das zur Modellierung des Erwerbs einfacher arithmetischer Fertigkeiten formulierte wissenschaftliche System neben empirisch auftretenden Fehlern auch unsinnige Fehler generiert. Tatsächlich zeigte sich beim Vergleich verschiedener Modellvarianten, daß mit einer zunehmenden Anzahl erklärter Fehler auch die Häufigkeit bisher nicht beobachteter und teilweise völlig unplausibler Varianten weit überproportional anstieg.

Modellexperimente eröffnen eine weitere Prüfmöglichkeit. Die Modellannahmen werden gezielt modifiziert, etwa indem ausgewählte Wissensbestände entfernt werden. Anschließend wird analysiert, ob sich die theoretisch vorhergesagten Verhaltensänderungen, etwa die Nichtlösbarkeit spezieller Aufgaben in den Modelldaten nachweisen lassen (vgl. Young & O’Shea, 1981).

Insgesamt kommt der Prüfung der Suffizienz bei Modellierungen mittels wissenschaftlicher Systeme eine besondere Rolle zu. Tatsächlich bietet dieser Typ von Modellierungen wie kaum ein anderer die Möglichkeit zu solchen Prü-

fungen. Aus dieser Perspektive wird verständlich, warum Vertreter dieser Forschungsrichtung anderen Ansätzen der Theorieformulierung, bei denen Sufizienz überhaupt nicht gegeben ist, sehr skeptisch gegenüber stehen.

#### **4.1.5 Vergleich von Modell- und Personendaten**

Der Vergleich von Modell- und Personendaten bleibt der wichtigste Weg zur Prüfung der empirischen Adäquatheit einer Modellierung. Wenden wir uns zunächst den Fällen zu, in denen ein direkter Vergleich der Modelldaten über Einzelverhalten mit den Daten einer oder mehrerer Personen erfolgt. Dies ist dann naheliegend, wenn in der Modellierung allgemeine **und** individuelle Komponenten berücksichtigt sind bzw. wenn anhand einer Einzelfallstudie das Verhalten einer einzigen Person rekonstruiert wurde. Zumeist liegt im letztgenannten Fall eine umfassende und dichte Datenerhebung bei einer über einen längeren Zeitraum hinweg beobachteten Person vor (vgl. etwa Simon & Simon, 1978). Dies ermöglicht den Vergleich des Datenprotokolls der Person mit den Aufzeichnungen der Modelldaten. Im Sinne einer Kreuzvalidierung ist es dabei erforderlich, nur einen Teil der Daten zur Modellierung heranzuziehen, um den anderen mit aus dem Modell generierten Prognosedaten vergleichen zu können.

Eine besondere Form einer individualisierten Modellierung stellt die anhand des Systems MULEDS bereits illustrierte kognitionswissenschaftliche qualitative Diagnostik dar (vgl. für ein klassisches Diagnosesystem DEBUGGY, Burton, 1982, und allgemein Spada & Reimann, 1988; Wenger 1987). Bei dieser Form der Diagnostik, für die etwa die Erfassung des Wissens der einzelnen Personen ein typisches Beispiel ist, liegt ein Rahmenmodell vor, das allgemeine, für alle Personen geltende Annahmen beinhaltet. Darüber hinaus werden im Diagnoseprozeß für jede einzelne Person individuelle Anteile festgelegt, beispielsweise die eine Person charakterisierenden korrekten und fehlerhaften Wissenskomponenten. Für jede einzelne Person resultiert ein individualisiertes Modell in Form eines lauffähigen wissensbasierten Systems. Worin besteht nun die empirische Prüfung der einzelnen Modelle und des Diagnosesystems als Ganzem? Anhand der empirischen Daten wird für jede einzelne Person ein Modell spezifiziert. Dabei wird geprüft, ob es möglich ist, anhand dieser Modelle die beobachteten Reaktionen der Personen etwa auf Aufgaben des untersuchten Gegenstandsbereichs zu rekonstruieren. Noch überzeugender ist die Prüfung, wenn sie im Sinne einer Kreuzvalidierung an einem neuen Satz von Daten derselben Personen erfolgt. Das einzelne Modell wird verworfen, wenn die mit seiner Hilfe generierten Daten nicht mit den Antworten der betreffenden Person übereinstimmen. Das gesamte Diagnosesystem ist zurückzuweisen, wenn es bei einer größeren Anzahl der Probanden versagt.

Schwieriger ist die Situation, wenn die Modellierung in einer prototypischen Form vorliegt, die keine individuenbezogene Spezifikationen erlaubt. Dann ist von vornherein damit zu rechnen, daß ein direkter Vergleich von Modell-daten mit den Daten einzelner Personen Abweichungen signalisiert. Wir haben ja bei Modelliermagazinen dieses Typs mit wissensbasierten Systemen die besondere Situation, daß eine allgemeine, auf den prototypischen Probanden bezogene Theorie so formuliert ist, daß sie Einzelverhaltensdaten generiert, die prinzipiell einem Vergleich mit den Daten beliebiger Personen unterzogen werden können. Welcher Weg kann in diesem Fall sinnvollerweise beschritten werden?

Eine empirische Geltungsprüfung ist beispielsweise durch Ableitung von Hypothesen über Daten anhand der Modellierung ganz im Sinne des klassischen, experimentell-statistischen Ansatzes möglich. Ausgehend vom Annahmengengefüge des Modells werden Bedingungen spezifiziert, die theoretisch interessante Effekte auf die Daten erwarten lassen. Unter diesen Bedingungen werden sowohl Modellläufe durchgeführt, als auch Personen getestet. Es wird geprüft, ob sich die in den Modelldaten niederschlagenden Bedingungeffekte auch in statistisch gesicherter Weise anhand der Personendaten feststellen lassen. Charakteristisch für diese Form der empirischen Modellprüfung ist die hohe Datenrestriktion, die ganz im Gegensatz zum Umgang mit Daten in der Phase der Modellentwicklung steht.

Ein Beispiel für eine derartige Modellprüfung wäre, bezogen auf das schon erwähnte Wissenserwerbsmodell KAGE (vgl. Tabelle 3), daß verschiedene Sequenzen von Lernaufgaben dargeboten werden. Sie werden so ausgewählt, daß sie bei Anwendung des Systems KAGE zu einem optimalen, einem suboptimalen und einem Lernergebnis mit fehlerhaftem Wissen führen. An Schülern wird überprüft, ob sich die anhand des Modells vorhergesagten Effekte der Lernbedingungen empirisch bestätigen lassen.

Selbstverständlich prüft ein experimentell-statistisches Vorgehen dieses Typs nicht das gesamte Annahmengengefüge eines wissensbasierten Prozeßmodells. Es stellt sich die Frage nach dem durch die jeweilige Bedingungsvariation erreichten Auflösungsgrad, die bei anders formulierten Theorien allerdings häufig vernachlässigt wird. Und es ist das Problem zu berücksichtigen, auf das Erdfelder und Bredenkamp (in diesem Band) hinweisen, wenn sie die fehlende Passung von Hypothesen über Einzelverhalten und statistischen Populationsaussagen anmerken.

Diese Probleme sind keineswegs spezifisch für die Überprüfung von kognitiven Modellierungen mit Hilfe eines experimentell-statistischen Vorgehens, sondern treffen in gleicher Weise auf die empirische Kontrolle jeder umfassenden Theorie zu. Es hängt vom theoretischen Verständnis und experimen-

tellen Geschick des einzelnen Forschers ab, inwieweit es ihm gelingt, auf experimentellem Wege wichtige Modellaspekte zu prüfen.

#### **4.1.6 Modellvergleichende Argumentation**

VanLehn et al. (1984) setzen sich mit der Frage auseinander, wie nach einer erfolgreichen Suffizienzprüfung einer kognitiven Modellierung Schritte in Richtung auf das Kriterium „Notwendigkeit“ gemacht werden können. Wie lassen sich Hinweise gewinnen, daß die formulierten Annahmen und postulierten Prozesse notwendig sind, um das interessierende Verhalten zu generieren. Sie schlagen dafür ein Verfahren vor, das sie **modellvergleichende Argumentation** (competitive argumentation) nennen. Die Annahmen einer computersimulierten Theorie werden mit denen anderer Theorien oder aber anderer Versionen der Theorie selbst verglichen. Beispielsweise gilt es zu zeigen, daß ein Designprinzip, das der Entwicklung eines wissensbasierten Systems zugrundeliegt, bestimmte Fakten zu erklären vermag, während Variationen dieses Prinzips, obwohl selbst nicht ohne empirischen Wert, in der einen oder anderen Weise mit Problemen belastet sind. Um zu belegen, daß eine bestimmte, in eine Modellierung eingeführte Bedingung notwendig ist, genügt es keineswegs zu zeigen, daß das Suffizienzkriterium erfüllt ist. Wenn nämlich zwei Theorien vorliegen, von denen die eine behauptet, daß Prinzip X hinreichend sei und eine andere, daß dies für ein mit X inkompatibles Prinzip Y gelte, dann versagt das Suffizienzkriterium als Richtschnur zur Entscheidung zwischen den beiden Modellierungen. Es muß vielmehr argumentiert werden, warum Prinzip X aus psychologischer Sicht besser ist als Prinzip Y. In dieser Art einer vergleichenden Argumentation sehen VanLehn et al. (1984) eine realistische Alternative zum letztlich unmöglichen Nachweis, daß eine Annahme notwendig ist. Vergleichende Argumentation sei ein Weg für eine sukzessive Annäherung an psychologisch valide Theorien. Des weiteren verweisen VanLehn et al. (1984) darauf, daß derartige Argumente, etwa empirische Befunde, die eine von mehreren alternativen Theorien belegen, häufig einen längeren Bestand haben, als die Theorien selbst, zu deren Stützung sie vorgelegt wurden.

Eine modellvergleichende Argumentation kann als Verallgemeinerung des bekannten Entscheidungsexperiments, bei dem zwischen zwei Theorien auf der Basis eines empirischen Befundes entschieden wird, angesehen werden. In dieselbe Kategorie fällt auch der allerdings formal wesentlich ausgefeiltere Vergleich von mathematisch-numerischen Modellen, wenn etwa bei probabilistischen Ansätzen ein Likelihood-Vergleich durchgeführt wird, um eine Aussage über die relative Geltung zweier konkurrierender Modelle treffen zu können.

## 4.2 Weitere Bewertungskriterien

Gadenne (in diesem Band) führt weitere Kriterien zur Theoriebewertung an: logische Konsistenz und Widerspruchsfreiheit, semantische Einheitlichkeit, Informationsgehalt und Tiefe von Theorien sowie Einfachheit. Wie sind anhand dieser Bewertungskriterien Modellierungen in Form wissensbasierter Systeme zu beurteilen? Selbstverständlich ist es schwierig, generelle Aussagen zu treffen. Die folgenden Feststellungen beschreiben deshalb den Grad an Kriterienerfüllung, der bei diesem Typ von Theorieformulierung bei einem adäquaten Vorgehen prinzipiell erreichbar ist.

Zur Frage der logischen Konsistenz und Widerspruchsfreiheit bieten lauffähige wissensbasierte Systeme über die Generierung repräsentativer Sätze von Modelldaten eine gute Prüfmöglichkeit. Die Simulation eröffnet einen Weg, Widersprüchen auf die Spur zu kommen. Vergleichbare Prüfmöglichkeiten weisen verbal formulierte Theorien nicht auf.

Die Frage nach der semantischen Einheitlichkeit, also etwa der Einheitlichkeit des verwendeten Begriffssystems ist schwer zu beantworten. Sicherlich erzwingt die Formalisierung durch die jeweilige Festlegung auf einen bestimmten Programmierstil und eine bestimmte Programmiersprache von vornherein eine gewisse Einheitlichkeit. Es ist aber nicht gewährleistet, daß tatsächlich nur semantisch einheitliche Begriffe in die Formalisierung eingehen.

Die Kriterien „Informationsgehalt“ und „Tiefe“ sind bei wissensbasierten Systemen in hohem Maße erfüllt. Die Notwendigkeit der Explizitheit und Detaillierung zwingt zu vielen und präzisen Aussagen. Es resultieren Theorien mit hohem Informationsgehalt und zugleich großer Tiefe. Es ist ja gerade das Anliegen, die „black box“ zwischen Ein- und Ausgabegrößen aufzubrechen und die Prozesse zu spezifizieren, die den Zusammenhang zwischen diesen Größen bedingen.

Einfachheit ist nicht die Stärke von Modellierungen mit Hilfe wissensbasierter Systeme. Das Ziel der Suffizienz und das damit verbundene Merkmal der Lauffähigkeit zwingen dazu, manche Annahmen über Details einzuführen, die nicht im Kern der Theorie stehen und über die möglicherweise keine hinreichende theoretische und empirische Evidenz vorliegt.

Die bisher genannten Theoriebewertungskriterien stammen weitgehend aus einer Zeit, in der der Stellenwert computersimulierter Theorien noch sehr gering war. Müssen sie nicht mit Blick auf diesen neuen Typ der Theorieformulierung modifiziert, präzisiert und vor allem ergänzt werden?

Ohlsson (1988) führt zwei Bewertungskriterien an, die aufs engste mit der Lauffähigkeit wissensbasierter Systeme und mit dem neuen Typ von Perfor-

manz, der Generierung modellbasierter Daten, zu tun haben. Es sind dies die schon angesprochene Suffizienz einer Modellierung und die Anwendbarkeit des wissensbasierten Systems. Das letztgenannte Kriterium hängt eng mit einem klassischen Theorieideal zusammen, nämlich der Frage der Verallgemeinerbarkeit der Ergebnisse. Es gewinnt allerdings im Kontext computersimulierter Theorien einen besonderen Stellenwert, da sich über die Generierung von Modelldaten neue Anwendungsmöglichkeiten ergeben. Beispiele sind etwa:

- die Ableitung präziser Verhaltensvorhersagen unter verschiedenen situativen Bedingungen,
- die Aufdeckung von kognitiven Strukturen und Prozessen, die zu fehlerhaften Reaktionen führen, und
- die Ableitung von Interventionsmaßnahmen mit in der Simulation erprobten Effekten.

Der letzte Punkt läßt sich an dem Konzept der ‚Pseudostudenten‘ von Van-Lehn (1991) verdeutlichen. Er nutzt Computermodellierungen des Wissenserwerbs, um verschiedene Instruktionsstrategien zu erproben. Aus der Simulation wird abgeleitet, welche Strategie den Lernzielen am besten dient. Erst dann erfolgen Felduntersuchungen. Computersimierte Theorien des Wissenserwerbs werden hier zum Testfeld für Instruktionsfragen. Die gezielt eingesetzte Empirie dient der Überprüfung der über die Modelldaten gewonnenen Erwartungen. Im Vergleich zu empirischen Interventionsstudien nach dem Versuch und Irrtum Prinzip ist dieses Vorgehen wesentlich ökonomischer und theoretisch befriedigender.

Tabelle 4: Vergleich formaler Modelle numerischer Art und nicht-numerischer Art

Inwieweit erfüllen formale Modelle numerischer Art (wie mathematische Lernmodelle) und nicht-numerischer Art (wie wissensbasierte Systeme) die bekannten Theoriebewertungskriterien (vgl. Gadenne, in diesem Band)?

Konnektionistische Modelle sind hier ausgeklammert. Auch zur **inhaltlichen** Frage der empirischen Adäquatheit einer Theorie werden keine Aussagen gemacht.

formale Modelle	logische Konsistenz	Informationsgehalt/Tiefe	Einfachheit	Suffizienz	Empirische Prüfbarkeit	Praktische Anwendbarkeit
numerischer Art	ja	eher gering	ja	eher nein	gut	eher gering
nicht-numerischer Art	ja	groß	nein	ja	eher schlecht	oft gegeben

Das zweite, nicht nur von Ohlsson (1988) ins Gespräch gebrachte Theoriebewertungskriterium ist das der Suffizienz. Sie ist Vorbedingung für empirische Adäquatheit. Das Streben nach suffizienter Modellierung hat vielfältige

Effekte auf den Forschungsprozeß. Ein noch nicht angesprochener Effekt ist die Notwendigkeit einer ganzheitlichen Betrachtungsweise im Sinne einer integrierten Behandlung aller Strukturen und Prozesse, die zur Erbringung einer Leistung erforderlich sind. Eine Beschränkung auf eine einzelne Komponente wäre unter diesem Gesichtspunkt nicht hinreichend.

Tabelle 4 faßt die Ausführungen zur Bewertung wissensbasierter Systeme nochmals zusammen, in dem formale Modelle numerischer und nicht-numerischer Art schematisch anhand wichtiger Theoriebewertungskriterien verglichen werden.

## 5. *Wissenschaftstheoretische Anmerkungen*

Ist es wissenschaftstheoretisch gesehen überhaupt begründbar, daß geistige Aktivitäten und intelligente Leistungen des Menschen mit technischen Mitteln untersucht und im Sinne einer maschinellen Rekonstruktion auf mechanische Weise nachgebildet werden?

Zur Versachlichung der Diskussion dieser Frage sollen zunächst zwei Thesen unterschieden werden:

These 1: Wissensbasierte Systeme sind ein nützliches und mächtiges Werkzeug zur Erforschung der menschlichen Kognition. Die Rekonstruktion kognitiver Prozesse in Form von Berechnungsvorgängen ist möglich und theoretisch sinnvoll. Befürworter dieser, nach einer Bezeichnung von Searle (1980) als schwache These **der KI** bekanntgewordenen Position verweisen typischerweise auf die präzisere Formulierung, den größeren Komplexitätsgrad und den höheren Informationsgehalt der zugrundeliegenden Theorien sowie auf die damit einhergehende Eröffnung vielfältiger Anwendungsperspektiven.

These 2: Wissensbasierte Systeme sind nicht nur ein Mittel zur Erforschung der menschlichen Kognition, sondern sie können darüber hinaus selbst als Träger kognitiver Prozesse angesehen werden. Kognition vollzieht sich in Prozessen der Informationsverarbeitung, d.h. kognitive Prozesse sind Berechnungsprozesse. Diese **starke** These **der KI** (Searle, 1980) besagt, daß im Unterschied etwa zur Simulation eines Binnensees mit Hilfe von mathematisch-numerischen Modellen, die natürlich nicht selbst ein Binnensee ist, die Simulation eines kognitiven Prozesses durch ein wissensbasiertes System selbst wiederum ein kognitiver Prozeß ist, da in diesem Fall das Simulierte und seine Simulation nach denselben Prinzipien der Symbolverarbeitung funktionieren. Oder noch provokanter formuliert: In wissensbasierten Systemen laufen geistige Prozesse ab. Solche Systeme besitzen Geist.

Gegenstand der folgenden Ausführungen ist die Frage nach dem wissenschaftstheoretischen Selbstverständnis der in den vorangegangenen Abschnitten dargestellten Methodologie. Ausgangspunkt ist der sog. Turing-Test, der einen pragmatischen, operationalen Zugang zu dem gesamten Problemkreis liefert. Anschließend wird mit der Position des Funktionalismus eine einflußreiche wissenschaftstheoretische Position zur Philosophie des Geistes in ihren Grundannahmen skizziert, bevor dann anhand des Gedankenexperiments von John Searle zum sog. chinesischen Zimmer die Position der starken KI-These kritisch diskutiert wird.

## 5.1 Der Turing-Test

Alan Turing beginnt seine berühmt gewordene, 1950 veröffentlichte Arbeit „Computing machinery and intelligence“ mit den folgenden Worten:

I propose to consider the question, ‘Can machines think?’. This should begin with definitions of the meaning of the terms ‘machine’ and ‘think’. . . .

(Turing, 1950, S. 433)

Im folgenden expliziert er den heute als **Turing-Test** bekannten Gedanken-gang. Dabei geht es um eine Operationalisierung der Frage, wann einem Computer die Fähigkeit zu denken zugeschrieben werden kann. Turing wählt einen pragmatischen Zugang zur Beantwortung: Ein Computer denkt, sofern er ununterscheidbar von einer Person handelt. Dies kann aufgrund eines Imitations-tests entschieden werden. Eine Person soll eine andere Person und einen Computer (bzw. ein Computerprogramm) befragen. Die gesamte Kommunikation (Fragen und Antworten) wird mit technischen Mitteln, ohne jeden Sichtkontakt durchgeführt. Die Aufgabe für die fragende Person besteht darin, die andere Person bzw. den Computer zu identifizieren, indem sie das Frage-Antwort-Spiel entsprechend gestaltet. Falls die fragende Person auch nach einer Reihe von Fragen die andere Person nicht eindeutig zu identifizieren vermag, gesteht man dem Computer zu, den Test erfolgreich bestanden zu haben.

Unabhängig von der Tatsache, daß auch heute noch kein Computersystem existiert, daß diesen Test erfolgreich bestanden hat, führten diese Überlegungen zu einer intensiven Diskussion der Frage, ob Menschen und Computer prinzipiell vergleichbar sind und ob der Turing-Test ein angemessenes Entscheidungsverfahren darstellt. Muß man einem Computer, der den Turing-Test besteht, **notwendigerweise** kognitive Fähigkeiten zuschreiben?

Ein in diesem Zusammenhang immer wieder geäußerter Einwand bezieht sich auf den behavioristisch verengten Vergleichsmaßstab: Nicht das Denken selbst (in Form interner, nicht beobachtbarer mentaler Vorgänge) wird thematisiert,

sondern es werden lediglich die Denkergebnisse in Form beobachtbaren (Antwort-) Verhaltens untersucht. Aber ist es andererseits nicht vernünftig anzunehmen, daß sich jede noch so geschickte Imitation durch zielgerichtetes Prüfen aufdecken läßt, falls ein künstliches System wie ein Computer sein Verhalten auf prinzipiell andere Weise hervorbrächte als ein Mensch? Was ist überhaupt der Status mentaler Vorgänge bzw. Zustände? Handelt es sich um mehr als bloße sprachliche Konventionen? Mit Fragen dieser Art setzt sich die Philosophie des Geistes im Allgemeinen und die Position des Funktionalismus im Besonderen auseinander. Sie sind auch für die Philosophie der Psychologie von grundlegender Bedeutung (vgl. für eine umfassende Darstellung verschiedener Positionen Carrier & Mittelstraß, 1988).

## 5.2 Der Funktionalismus

Der philosophische **Funktionalismus** versucht eine Antwort auf die Frage zu geben, wie mentalistische bzw. intentionale Begriffe wie Wissen (im Sinne von Überzeugungen), Erwartungen, Ziele oder Motive interpretiert werden können (vgl. Carrier & Mittelstraß, 1988; Gadenne, Theorien, in diesem Band). Worauf beziehen sich derartige Begriffe? Welchen Status besitzt eine Aussage der Art „Eine Person befindet sich in einem bestimmten mentalen Zustand, etwa sie hat das Ziel, ein gegebenes Problem zu lösen?“

Der Funktionalismus ist durch die Auffassung gekennzeichnet, daß mentale Zustände ihrer logischen Natur nach von materiellen Zuständen als verschieden angesehen werden müssen. Mentale Zustände können nicht sinnvoll durch Eigenschaften ihrer materiellen Realisierung, also etwa als Gehirnzustände definiert werden, sondern nur funktional. Sie sind charakterisiert durch ihre kausalen Beziehungen zu anderen mentalen Zuständen, zu äußeren Bedingungen und zum Verhalten. Entscheidend ist ihr kausales Potential, durch das sie in der Informationsverarbeitung wirksam werden können. Ein abstrakter funktionaler Zustand ist auf multiple, verschiedenartige Weise materiell realisierbar. Daraus folgt umgekehrt, daß für die Beschreibung der funktionalen Organisation eines informationsverarbeitenden Systems, verstanden als die Gesamtheit seiner funktionalen Zustände und ihrer Relationen, die materielle Realisierung nicht bekannt sein muß.

Angenommen, wir sind Zeugen eines schweren Verkehrsunfalls und beobachten, daß ein anderer Zeuge daraufhin zur nächsten Telefonzelle eilt. Dieses Verhalten kann offensichtlich funktional erklärt werden, wobei die resultierende Erklärung auch zur Verhaltensvorhersage genutzt werden kann. Würde es Sinn machen, dieses Verhalten neurobiologisch beschreiben und erklären zu wollen? Es ist eine der Grundthesen der Kognitionswissenschaft, daß die Erklärung menschlichen Verhaltens durch Rückgriff auf mentalistische bzw.

intentionale Begriffe das erfolgreichste verfügbare Modell für menschliche Verhaltensklärungen darstellt. Damit geht die Auffassung einher, daß mit Hilfe funktionaler Erklärungen Regelförmigkeiten formuliert werden können, die mit anderen Kategorien von Begriffen nicht adäquat faßbar sind.

Welche Implikationen ergeben sich aus einer derartigen Betrachtungsweise? Menschen und in geeigneter Weise programmierte Computer werden als informationsverarbeitende Systeme angesehen, die eine zwar materiell unterschiedlich realisierte, aber dennoch prinzipiell vergleichbare funktionale Organisation aufweisen können. Aus dieser Vergleichbarkeit wird die Gültigkeit der starken These der KI gefolgert. Der Funktionalismus geht davon aus, daß die Prinzipien der funktionalen Organisation eines wissensbasierten Systems im Prinzip denen des menschlichen kognitiven Apparats gleichen: Maschinen sind grundsätzlich zu denselben kognitiven Leistungen wie der Mensch in der Lage (vgl. auch die auf Newell und Simon zurückgehende **Hypothese der Symbolsysteme**; Newell & Simon, 1976).

Für die kognitionspsychologische Theoriebildung hat diese Betrachtung, auch wenn man dem Funktionalismus in seiner extremen Form nicht folgt, zwei Konsequenzen. Erstens, intentionale Begriffe wie Wissen und Ziele sind wissenschaftlich sinnvolle, nicht eliminierbare Konzepte, und es ist methodologisch gerechtfertigt, sie bei der Beschreibung und Erklärung kognitiver Strukturen und Prozesse zu verwenden. Dabei werden derartigen mentalen Inhalten in Analogie zur Arbeitsweise eines wissensbasierten Systems sowohl syntaktisch-kausale als auch semantisch-logische Eigenschaften zugewiesen (vgl. die Hypothese der Wissensrepräsentation in Abschnitt 3.1.3). Zweitens sind alle postulierten Inhalte so zu präzisieren, daß sie als Programm einer Turing-Maschine, also in algorithmischer Weise formulierbar sind.

Zusammenfassend sind mentale Zustände aus der Sicht des Funktionalismus semantisch interpretierbare Größen, die bestimmte Inhalte ausdrücken und insofern semantische Attribute wie Wahrheitswerte oder Referenz aufweisen. Daneben kommen mentalen Zuständen aber auch syntaktische oder formale Eigenschaften zu, die alle diejenigen Merkmale umfassen, die von ihrer Bedeutung unabhängig sind. Damit korrespondiert, daß **kognitive Aktivitäten** als regelhafte Prozesse **über semantisch interpretierten symbolischen Repräsentationen** angesehen werden können, wie dies Pylyshyn (1987) in seinem „**representational metapostulate**“ explizit formuliert hat:

Computation, information processing, and rule-governed behavior all depend on the existence of physically instantiated codes or symbols that refer to or represent things and properties extrinsic to the behaving system. In all these cases the behavior of the systems in questions (be they minds, computers or social systems) are explained not in terms of intrinsic properties of the system itself but in terms of rules and processes that operate on representations of extrinsic things. **Cognition, in other words, is ex-**

**plained in terms of regularities holding over semantically interpreted symbolic representations, . . .** (Hervorhebung von den Verfassern).

(Pylyshyn, 1987, S. 133)

### 5.3 Das chinesische Zimmer

Die starke These der KI bzw. die These einer prinzipiellen Vergleichbarkeit menschlichen und maschinellen Verstehens ist aus verschiedenen Positionen heraus angegriffen worden (vgl. etwa Dreyfus, 1979; Winograd & Flores, 1987; Penrose, 1989). Die Position des vielleicht bekanntesten Kritikers, John Searle, soll in diesem Abschnitt skizziert werden. Dabei wird anknüpfend an die Überlegungen zum Turing Test insbesondere die Frage diskutiert, ob aus dem Verhalten eines Systems dessen kognitive Fähigkeiten erschließbar sind bzw. anders formuliert, ob ‚intelligentes‘ Handeln Verstehen voraussetzt.

Unter dem Begriff **chinesisches** Zimmer ist ein Gedankenexperiment berühmt geworden, das auf Searle (1980) zurückgeht. Ein modernes KI-Sprachverarbeitungssystem besitzt zweifellos eine erstaunliche Leistungsfähigkeit. Es liefert gute Fragen und vernünftige Antworten. Was aber versteht ein „sprachverstehendes“ System wirklich? Ist es nicht so, daß lediglich eine rein syntaktische Verknüpfung gewisser Zeichenreihen, die dem System eingegeben werden, mit gewissen anderen Zeichenreihen, die vom System ausgegeben werden, erfolgt, ohne daß dabei die Bedeutung der Zeichen in irgendeiner Weise ‚verstanden‘ wird? Zur Diskussion dieser Frage entwirft Searle das folgende Szenario:

Eine Person, etwa Searle selbst, befinde sich in einem abgeschlossenen Raum, vor sich einen Stapel in chinesischer Sprache verfaßten Textes, von dem er nicht ein Wort versteht, ja noch nicht einmal weiß, daß es sich um chinesische Schriftzeichen handelt. In diesen Raum wird nun zunächst von außen ein zweiter Stapel chinesischen Textes hereingereicht, zusammen mit einer Menge von in englischer Sprache formulierten Regeln, mit deren Hilfe die Elemente des ersten und des zweiten Stapels auf rein mechanische Weise zueinander in Beziehung gesetzt werden können (etwa ausschließlich aufgrund der Berücksichtigung ihrer äußerlichen Form). Schließlich werde noch ein dritter Stapel chinesischen Textes, zusammen mit einer Menge von englischsprachlichen Anweisungen zur Verfügung gestellt, die es Searle ermöglichen, Elemente des dritten Stapels mit Elementen aus den beiden anderen Stapeln in Beziehung zu bringen und für bestimmte Zeichen aus dem dritten Stapel mechanisch bestimmbare chinesische Zeichen aus den beiden anderen Stapeln zurückzugeben.

Bezugnehmend auf das Skriptkonzept bzw. die hierauf beruhenden Dialogsysteme von Roger Schank (vgl. Schank & Abelson, 1977), die den Anspruch erheben, Sprache „verstehen“ zu können (im Unterschied zum bloßen Erwecken eines entsprechenden Anscheins), führt Searle weiter aus:

Unknown to me, the people who are giving me all of these symbols call the first batch a “script”, they call the second batch a “story”, and they call the third batch “questions”. Furthermore, they call the symbols I gave them back in response to the third batch “answers to the questions” and the set of rules in English that they gave me, they call “the program”. (S.418)

Werden nun Searle in dieser Situation chinesische „Fragen“ hereingereicht, dann können zwar für einen externen Beobachter die chinesischen Zeichen, die Searle als „Antworten“ zurückgibt, durchaus als „sinnvolle“ Antworten erscheinen, vorausgesetzt der externe Beobachter ist der chinesischen Sprache mächtig. Searle aber, der die Fragen mechanisch mit Hilfe der englischen Regeln und Anweisungen und ausschließlich aufgrund äußerlicher („syntaktischer“) Merkmale beantwortet hat, versteht nach wie vor offenbar kein einziges Wort Chinesisch:

It seems to me quite obvious in the example that I do not understand a word of Chinese stories. I have inputs and outputs that are indistinguishable from those of the native Chinese speaker, and I can have any formal program you like, but I still understand nothing. For the same reasons, Schank's computer understands nothing of any stories, whether in Chinese, English, or whatever, since in the Chinese case the computer is me, and in cases where the computer is not me, the computer has nothing more than I have in the case where I understand nothing. (S.418)

Nehmen wir zur Komplikation der Situation weiter an, daß man dem eingeschlossenen Searle nicht nur chinesische, sondern auch englische Fragen gibt, die Searle als „native speaker“ zweifellos so beantworten könnte, daß seine Antworten für einen externen Beobachter als die Antworten einer die englische Sprache „verstehenden“ Person beurteilt werden. Gibt es zwischen beiden Situationen (Verstehen chinesischer vs. englischer Sprache) einen Unterschied und wenn ja, woran ist er zu erkennen?

Searle ist der Meinung, daß es einen prinzipiellen Unterschied zwischen seinem Verstehen als Mensch und dem Verstehen eines Computers gibt. Menschen verfügen danach über die Fähigkeit, Bedeutung unmittelbar zu erfassen, wofür er den Begriff der **intrinsic Intentionalität** einführt. Menschen wissen, was sie meinen - meint Searle - niemals jedoch ein Computer, da seine Arbeitsweise auf rein syntaktisch-formalen Operationen beruht, die keine Semantik beinhalten. Er begründet seine Argumentation im Kern damit, daß die Eigenschaft der intrinsic Intentionalität an die besondere biologische Ausstattung des Menschen, insbesondere an sein Gehirn, gebunden ist. Diese spezielle materielle, genauer gesagt neurobiologische Bedingtheit mentaler Zu-

stände impliziert für Searle, daß die starke These der KI als unzutreffend zurückzuweisen ist.

Diese Argumentation Searles ist selbst wiederum keineswegs unumstritten. Befürworter des Funktionalismus begegnen ihr etwa mit dem Einwand, daß die Zuschreibung von Verstehen nicht von einer Kenntnis der internen Zustände bzw. Gehirnzustände einer Person abhängen könne. Wovon aber dann? Zu dieser Frage kann Searle keine konstruktive Perspektive aufzeigen. Für ihn steht der Gedanke im Vordergrund, daß ‚richtiges‘ Handeln kein hinreichendes Kriterium für Verstehen ist. Aber welche empirischen Grundlagen für die Zuschreibung kognitiver Fähigkeiten blieben dann für die Psychologie? Damit sind wir aber wieder bei Fragen, die seit Turing im Mittelpunkt der wissenschaftstheoretischen Diskussion stehen.

## 6. Epilog

Was ist aber nun unsere eigene kognitionspsychologische Perspektive auf die zuvor aufgeworfenen wissenschaftstheoretischen Fragen?

Für uns sind wissensbasierte Systeme - entsprechend der schwachen These der KI - ein wichtiges Werkzeug zur Erforschung der menschlichen Kognition. Aber auch die starke These wird von uns nicht als abwegig beurteilt. Dabei ist sorgfältig zwischen den kognitiven Prozessen selbst und ihrem jeweiligen Träger zu unterscheiden. Der Mensch **hat** Geist und er **ist** intelligent, während die Maschine agiert, als **ob** sie Geist hätte, als **ob** sie intelligent sei. Die kognitiven Prozesse selbst können je nach ihren Eigenschaften - aber unabhängig von der Form ihrer materiellen Realisierung und damit von ihrem Träger - als mehr oder minder flexible, intelligente usw. Prozesse bezeichnet werden. Das Modellerte und seine Modellierung funktionieren nach denselben Prinzipien der Informationsverarbeitung. Die Simulation eines Denkprozesses ist ein Denkprozeß, simulierte Wahrnehmung ist Wahrnehmung. Damit ist aber noch nichts darüber ausgesagt, ob menschliches Denken und Wahrnehmen nicht über den informativisch rekonstruierbaren Anteil hinaus durch weitere Anteile geprägt sind.

Diskussionen zur schwachen und starken These der KI übersehen auch häufig, daß man noch weit davon entfernt ist, menschliche Kognition so zu modellieren, daß die Rekonstruktion den Turing-Test bestehen könnte, oder Herrn Searle im chinesischen Zimmer dazu befähigen wurde, einen intelligenten, wenn auch vielleicht unverstandenen Dialog zu führen. Tatsächlich ist dieses Ziel noch fern. Aber wir glauben, daß es eine lohnende Forschungsperspektive aufzeigt.

Weiter oben haben wir darauf hingewiesen, daß für die Beschreibung der funktionalen Organisation eines informationsverarbeitenden Systems die materielle Realisierung nicht bekannt sein muß. Dies darf nicht dahingehend mißverstanden werden, daß die materielle Realisierung unerheblich sei. Sie legt den Raum der Möglichkeiten fest. Bei Ausfall bestimmter Komponenten auf der materiellen Ebene ergeben sich u.U. dramatische Konsequenzen für die funktionale Ebene.

Kehren wir aber zur Frage der Modellierung kognitiver Strukturen und Prozesse mit Hilfe wissensbasierter Systeme zurück. Wir sind der Auffassung, daß wissensbasierte Systeme durch eine Reihe von Eigenschaften charakterisiert sind, die sie zu einer ausgezeichneten Kategorie von Modellen machen, um unser Verständnis kognitiver Vorgänge beim Menschen zu verbessern. Sie erlauben eine Form der Theoriebildung, die durch grundlegende wissenschaftstheoretische Ideale gekennzeichnet ist. Darüber hinaus liegt wissensbasierten Systemen ein äußerst anspruchsvoller Erklärungs-begriff zugrunde. Wissensbestände definieren ein Kompetenzmodell im Sinne eines Potentials, das verhaltenswirksam werden kann, aber nicht muß. Ihre Aktivierung im Prozeß der Auswahl und Anwendung setzt die Spezifizierung generativer Mechanismen voraus. Die in Frage stehenden Performanzphänomene werden nicht nur oberflächlich beschrieben, sondern in einer konstruktiven Weise (re-)generiert. Die Eigenschaft der Suffizienz von Computermodellen liefert zugleich eine Art von **hinreichendem Existenzbeweis**, insofern das postulierte Modell sich als hinreichend zur Erzeugung der von ihm zu erklärenden Phänomene erwiesen hat.

Wissensbasierte Systeme führen zu einer **inkrementellen Theorieentwicklung**. Als generative Prozeßmodelle ermöglichen sie eine quasi-experimentelle Methodologie, insofern Hypothesen aufgestellt und ihre Implikationen als Simulationen unmittelbar im Modell überprüft werden können. Durch die Möglichkeit einer differenzierten Analyse der Leistungsfähigkeit, aber auch der Grenzen existierender Computersysteme sowie durch die Verwendung einer präzisen Sprache wird eine systematische Nutzung von bekannten Ergebnissen und ihre gezielte Weiterführung möglich. Neue Methoden wie beispielsweise die modellvergleichende Argumentation wurden entwickelt bzw. sind künftig zu entwickeln, um den damit einhergehenden Anforderungen im Prozeß der Theorieentwicklung und -testung gerecht werden zu können.

Wissensbasierte Systeme sind dann Teil (kognitions-) psychologischer Theoriebildung, wenn ein valides Abbild mentaler Strukturen und Prozesse angestrebt wird. Sind aber nur valide Modelle von Interesse, oder verdienen auch solche Entwicklungen unsere Aufmerksamkeit, bei denen der Mensch nicht explizit als Modell und Maßstab herangezogen wird und der ingenieurmäßige Aspekt der Modellierung im Vordergrund steht? Unsere Antwort auf diese

Frage ist durchaus positiv. Die Unterscheidung zwischen einer ingenieurmäßigen und einer psychologischen Betrachtung intelligenter Leistungen kann im Forschungsalltag kaum durchgehalten werden. Viele Realisierungen wissensbasierter Systeme orientieren sich in der einen oder anderen Form am Menschen und jede psychologisch motivierte Modellierung mit Hilfe wissensbasierter Systeme greift notwendigerweise auf in der Informatik entwickelte Methoden zurück. Ein Beispiel für die psychologische Fruchtbarkeit von Vorstellungen, für die nur selten der Anspruch empirischer Adäquatheit erhoben wird, ist der Bereich des maschinellen Lernens (vgl. etwa Kodratoff, 1988, oder das von Carbonell herausgegebene Sonderheft der Zeitschrift *Artificial Intelligence*, 1989). Diese Arbeiten haben in vielfältiger Weise psychologische Forschungen angeregt, die sich auf der Basis der in diesem Zusammenhang entwickelten Konzepte in neuer Weise mit Fragen menschlichen Lernens beschäftigen.

Diesen Argumenten für eine Modellierung kognitiver Vorgänge mit Hilfe wissensbasierter Systeme stehen aber auch eine Reihe von Problemen gegenüber. An erster Stelle ist das Spannungsfeld zwischen dem Auflösungsgrad, dem Aufwand einer Modellierung und ihrem Geltungsbereich zu nennen: Je feiner der Auflösungsgrad der Betrachtung, desto aufwendiger werden die Modelle, bei gleichzeitig eher eingeschränktem Geltungsbereich. Darüber hinaus hat die Anwendung von Methoden und Modellvorstellungen aus dem Computerbereich in fast schon paradoxer Weise dazu geführt, daß wir zunehmend genauer wissen, in welcher Weise menschliches Denken durch diese Modelle gerade nicht rekonstruiert werden kann. Obwohl wir über die Funktionsweise des informationsverarbeitenden Systems Mensch in vielen Bereichen nur eine begrenzte Anzahl positiver Aussagen machen können, so sind wir doch häufig sicher, daß es so nicht sein kann, wie es in einem speziellen wissensbasierten System postuliert wird, etwa aufgrund unrealistischer Annahmen zur Speicherkapazität.

Weiter haben die Forschungen der letzten vierzig Jahre deutlich gemacht, daß eine ausschließlich logisch-rationale Sicht von Kognition vielen grundlegenden Aspekten menschlichen Denkens und Handelns nicht gerecht wird. Auch ist es kein Zufall, daß gerade das sog. Alltagswissen mit all seinen Lücken, Unschärfen, Inkonsistenzen und Faustregeln sich einer Formalisierung bislang weitgehend entzieht, während für viele sog. ‚höhere‘ kognitive Funktionen gut bewährte Vorstellungen entwickelt werden konnten. In der Forschung zur Künstlichen Intelligenz ist man heute eher in der Lage, Schachprogramme mit der Spielstärke von Großmeistern zu entwickeln als ein halbwegs akzeptables Urlaubsberatungssystem. Kognitionspsychologisch läßt sich eher der über viel quantitatives Formelwissen verfügende physikalische Experte modellieren als ein Physikstudent, der mühsam damit kämpft, sein konzeptuelles Alltagswissen mit dem naturwissenschaftlichen Fachwissen zu integrieren. Und es bedarf

noch eines weiten Weges, einen Roboter zu entwickeln, der wohlbehalten eine Straße überquert, die gegenüberliegende Wirtschaft betritt, dort ein Glas Bier bestellt und eine Unterhaltung beginnt - vielleicht sollten wir letzteres aber auch uns vorbehalten.

## **Literatur**

- Aebli, H. (1980). **Denken: Das Ordnen des Tuns** (Band 1: Kognitive Aspekte der Handlungstheorie). Stuttgart: Klett-Cotta.
- Allen, J. (1987). **Understanding natural language**. Menlo Park, CA: Benjamin/Cummings.
- Anderson, J.R. (1982). Acquisition of cognitive skill. **Psychological Review**, 89, 369-406.
- Anderson, J. R. (1983). **The architecture of cognition**. Cambridge, MA: Harvard University Press.
- Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. **Psychological Review**, 94, 192-210.
- Anderson, J. R. (1990). **The adaptive** character of thought. Hillsdale, NJ: Erlbaum.
- Anzai, Y. & Simon, H. A. (1979). The theory of learning by doing. **Psychological Review**, 86, 124-140.
- Berwick, R. C. (1985). **The acquisition of syntactic knowledge**. Cambridge, MA: MIT Press.
- Brachman, R. & Levesque, H. (Eds.). (1985). **Readings in knowledge representation**. San Mateo, CA: Morgan Kaufmann.
- Brauer, W. (1990). Grenzen maschineller Berechenbarkeit. **Informatik-Spektrum**, 13, 61-70.
- Brown, J.S. & VanLehn, K. (1980). Repair theory: A generative theory of bugs in procedural skills. **Cognitive Science**, 4, 379-427.
- Burton, R.R. (1982). Diagnosing bugs in a simple procedural skill. In D. Sieeman & J. S. Brown (Eds.), **Intelligent tutoring systems** (pp. 157-183). New York, NY: Academic Press.
- Carbonell, J. G. (Ed.). (1990). Machine learning. **Artificial Intelligence**, 40.
- Carrier, M. & Mittelstraß, J. (1989). **Geist, Gehirn, Verhalten: Das Leib-Seele-Problem und die Philosophie der Psychologie**. Berlin: De Gruyter.
- Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, P. & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. **Cognitive Science**, 13, 145-182.
- Chomsky, N. (1963). Formal properties of grammars. In R.D. Luce, R.R. Bush & E. Galanter (Eds.), **Handbook of mathematical psychology**. Volume II (pp. 323-418). New York, NY: Wiley.
- Church, A. (1936). An unsolvable problem for elementary number theory. **American Journal of Mathematics**, 58, 345-363.

- Collins, A. M. & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407-428.
- Collins, A. M. & Smith, E. E. (Eds.). (1988). *Readings in cognitive science*. San Mateo, CA: Morgan Kaufmann.
- Daiser, W. (1984). **Künstliche Intelligenz** Forschung **und ihre epistemologischen Grundlagen**. Frankfurt/Main: Lang.
- Davis, E. (1990). **Representations of commonsense knowledge**. San Mateo, CA: Morgan Kaufmann.
- Deppe, W. (1977). **Formale Modelle in der Psychologie**. Stuttgart: Kohlhammer.
- Dörner, D. (1989). Die kleinen grünen Schildkröten und die Methoden der experimentellen Psychologie. *Sprache & Kognition*, 8, 86-97.
- Dreyfus, H. (1979). **What computers can't do: The limits of artificial intelligence**. New York, NY: Harper and Row. (Dt. Was Computer nicht können: Die Grenzen künstlicher Intelligenz. Königstein/Ts.: Athenäum, 1985).
- Gardner, H. (1985). **The mind's new science**. New York, NY: Basic Books (Dt. Dem Denken auf der Spur. Stuttgart: Klett-Cotta, 1989).
- Genesereth, M. R. & Nilsson, N.J. (1987). **Logical foundations of artificial intelligence**. Los Altos, CA: Morgan Kaufmann. (Dt. Logische Grundlagen der Künstlichen Intelligenz. Braunschweig: Vieweg, 1989).
- Gigerenzer, G. (1981). **Messung und Modellbildung in der Psychologie**. München: Reinhardt.
- Gigerenzer, G. (1988). Woher kommen Theorien über kognitive Prozesse? *Psychologische Rundschau*, 39, 91-100.
- Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, 98, 254-267.
- Gödel, K. (1931). über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme. *Monatshefte für Mathematik und Physik*, 38, 173-198.
- Herrmann, T. (1982). Über begriffliche Schwächen kognitivistischer Kognitionstheorien: Begriffsinflation und Akteur-System-Kontamination. *Zeitschrift für Sprache und Kognition*, 1, 3-14.
- Herrmann, T. (1988). Mentale Repräsentation - ein erläuterungsbedürftiger Begriff. *Sprache & Kognition*, 7, 162-175.
- Hildreth, E. C. & Ulkman, S. (1989). The computational study of vision. In M. I. Posner (Ed.), **Foundations of cognitive science** (pp. 581-630). Cambridge, MA: MIT Press.
- Holland, J., Holyoak, K., Nisbett, R. E. & Thagard, P. (1986). **Induction: Processes of inference, learning and discovery**. Cambridge, MA: The MIT Press.
- Hopcroft, J. E. & Ullman, J.D. (1979). **Introduction to automata theory, languages, and computation**. Reading, MA: Addison-Wesley.
- Johnson-Laird, P. N. (1983). **Mental models**. Cambridge: Cambridge University Press.
- Johnson-Laird, P.N. (1988). **The Computer and the mind**. Cambridge, MA: Harvard University Press.
- Kleene, S.C. (1936). General recursive functions of natural numbers. *Mathematical Annals*, 112, 727-742.

- Klix, F. (1988). Gedächtnis und Wissen. In H. Mandl & H. Spada (Hg.), *Wissenspsychologie (S. 19-54)*. München: Psychologie Verlags Union.
- Kodratoff, Y. (1988). *Introduction to machine learning*. London: Pitman.
- Krämer, S. (1988). *Symbolische Maschinen: Die Idee der Formalisierung in geschichtlichem Abriß*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Laird, J. E., Newell, A. & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence*, **33**, 1-64.
- Luce, R.D., Bush, R.R. & Galanter, E. (Eds.). (1963). *Handbook of mathematical psychology. Volume I-III*. New York, NY: Wiley.
- Malpas, J. (1987). *Prolog: A relational language and its applications*. Englewood Cliffs, NJ: Prentice-Hall.
- Markov, A. A. (1954). *Theory of algorithms*. Moskau: National Academy of Science.
- Marr, D. (1982). *Vision*. New York, NY: Freeman.
- McClelland, J. A. & Rumelhart, D. E. (Eds.). (1986). Parallel *distributed processing: Explorations in the microstructure of cognition. Volume II: Psychological and biological models*. Cambridge, MA: The MIT Press.
- McClelland, J. A. & Rumelhart, D. E. (1988). *Explorations in parallel distributed processing: A handbook of models, programs and exercises*. Cambridge, MA: The MIT Press.
- Minsky, M. (1967). *Computation: Finite and infinite machines*. Englewood Cliffs, NJ: Prentice-Hall.
- Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 211-277). New York, NY: McGraw-Hill.
- Möbus, C. (1988). Zur Modellierung kognitiver Prozesse mit daten- bzw. zielorientierten Regelsystemen. In H. Mandl & H. Spada (Hrsg.), *Wissenspsychologie (S. 423-468)*. München: Psychologie Verlags Union.
- Moravcsik, J. M. (1980). Chomsky's radical break with modern tradition (A commentary on Chomsky's Rules and Representations). *The Behavioral and Brain Sciences*, **3**, 28-29.
- Neches, R. (1982). Simulation systems for cognitive psychology. *Behavior Research Methods & Instrumentation*, **14**, 77-91.
- Newell, A. (1980). Physical symbol systems. *Cognitive Science*, **4**, 135-183.
- Newell, A. (1982). The knowledge level. *Artificial Intelligence*, **18**, 87-127.
- Newell, A.** (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Newell, A., Rosenbloom, P.S. & Laird, J.E. (1989). Symbolic architectures for cognition. In M.I. Posner (Ed.), *Foundations of cognitive science* (pp.93-132). Cambridge, MA: The MIT Press.
- Newell, A., Shaw, J. C. & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review*, **65**, 151-166.
- Newell, A. & Simon, H.A. (1963). GPS, a program that simulates human thought. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought* (pp. 279-293). New York, NY Mc Graw-Hill.

- Newell, A. & Simon, H.A. (1972). **Human problem solving**. Englewood Cliffs, NJ: Prentice-Hall.
- Newell, A. & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. **Communications of the ACM**, 19, 113-126.
- Ohlsson, S. (1988). Computer Simulation and its impact on educational research and practice. **International Journal of Educational Research**, 12, 5-34.
- Opwis, K. (1992). **Kognitive Modellierung: Zur Verwendung wissensbasierter Systeme in der psychologischen Theoriebildung**. Bern: Huber.
- Opwis, K. & Lüer, G. (in Druck). Modelle der Repräsentation von Wissen. In D. Albert & K.-H. Stapf (Hrsg.), **Enzyklopädie der Psychologie. Gedächtnispsychologie: Erwerb, Nutzung und Speicherung von Information**. Göttingen: Hogrefe.
- Osherson, D. N. & Lasniak, H. (Eds.). (1990). **An invitation to cognitive science. Volume 1: Language**. Cambridge, MA: MIT Press.
- Osherson, D. N. & Kosslyn, S.M. & Hollerbach, J.M. (Eds.). (1990). **An invitation to cognitive science. Volume 2: Visual cognition and action**. Cambridge, MA: MIT Press.
- Osherson, D. N. & Smith, E.E. (Eds.). (1990). **An invitation to cognitive science. Volume 3: Thinking**. Cambridge, MA: MIT Press.
- Penrose, R. (1989). **The emperor's new mind: Concerning computers, minds, and the laws of physics**. Oxford: Oxford University Press. (Dt. Computerdenken: Die Debatte um künstliche Intelligenz, Bewußtsein und die Gesetze der Physik. Heidelberg: Spektrum der Wissenschaft, 1991).
- Pinker, S. (1989). **Learnability and cognition: The acquisition of argument structure**. Cambridge, MA: MIT Press.
- Plötzner, R. & Spada, H. (1992). Analysis-based learning on multiple levels of mental domain representation. In E. De Corte, M. Linn, H. Mandl, & L. Verschaffel (Hrsg.), **Computer-based learning environments and problem solving** (pp. 103-127). Berlin: Springer.
- Plötzner, R., Spada, H., Stumpf, M. & Opwis, K. (1991). Learning qualitative and quantitative reasoning in a microworld for elastic impacts. **European Journal of Psychology and Education**, 4, 501-516.
- Posner, M.I. (Ed.). (1989). **Foundations of cognitive science**. Cambridge, MA: MIT Press.
- Post, E. L. (1943). Formal reductions of the general combinatorial decision problem. **American Journal of Mathematics**, 65, 197-215.
- Polyshyn, Z. W. (1985). **Computation and cognition**. Cambridge, MA: The MIT Press.
- Polyshyn, Z. W. (1987). Cognitive Science. In S. C. Shapiro (Ed.), **Encyclopedia of artificial intelligence, Volume 1** (pp. 120-124). New York, NY: Wiley.
- Polyshyn, Z. W. (1989). Computing in cognitive science. In M. I. Posner (Ed.), **Foundations of cognitive science** (pp.51-91). Cambridge, MA: The MIT Press.
- Quillian, M. R. (1968). Semantic memory. In M. Minsky (Ed.), **Semantic information processing** (pp.227-270). Cambridge, MA: The MIT Press.

- Reimann, P. & Schult, T. J. (1991). Modeling example elaboration strategies. In ***Proceedings of the First International Conference on the Learning Sciences, Evanston, IL***.
- Rogers, H. (1987). ***Theory of recursive functions and effective computability***. Cambridge, MA: The MIT Press.
- Rumelhart, D. E. & McClelland, J. A. (Eds.). (1986). ***Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations***. Cambridge, MA: The MIT Press.
- Schank, R. C. & Abelson, R.P. (1977). ***Scripts, plans, goals and understanding***. Hillsdale, NJ: Erlbaum.
- Scheffe, P. (1986). ***Künstliche Intelligenz - überblick und Grundlagen***. Mannheim: B. I. Wissenschaftsverlag.
- Searle, J. (1980). Minds, brains and programs. ***The Behavioral and Brain Sciences*, 3**, 417-457.
- Selz, O. (1913). über die Gesetze des geordneten Denkverlaufs**. Stuttgart: Speman.
- Simon, D.P. & Simon, (1978). Individual differences in solving physics problems. In R. Siegler (Ed.), ***Children's thinking: What develops*** (pp.325-348)? Hillsdale, NJ: Lawrence Erlbaum.
- Smith, B. C. (1982). ***Reflection and semantics in a procedural language*** (Technical Report). Cambridge, MA: The Massachusetts Institute of Technology.
- Smolensky, P. (1988). On the proper treatment of connectionism. ***Behavioral and Brain Sciences*, 11**, 1-74.
- Spada, H. & Reimann, P. (1988). Wissensdiagnostik auf kognitionswissenschaftlicher Basis. ***Zeitschrift für Differentielle und Diagnostische Psychologie*, 9**, 183-192.
- Stachowiak, H. (1973). ***Allgemeine Modelltheorie***. Berlin: Springer.
- Sterling, L. & Shapiro, E. (1986). ***The art of Prolog: Advanced programming techniques***. Cambridge, MA: MIT Press. (Dt. Prolog: Fortgeschrittene Programmier-techniken. Bonn: Addison-Wesley, 1988).
- Stetter, F. (1988). ***Grundbegriffe der theoretischen Informatik***. Berlin: Springer.
- Strube, (1990). Neokonnektionismus: Eine neue Basis für die Theorie und Modellierung menschlicher Kognition? ***Psychologische Rundschau*, 41**, 129-143.
- Tack, W. (1987). Ziele und Methoden der Wissensrepräsentation. ***Sprache & Kognition*, 3**, 150-163.
- Turing, A. (1936). On computable numbers with an application to the Entscheidungsproblem. ***Proceedings London Math. Society*, 42**, 230-265. A correction, *ibid.*, 43, 544-546.
- Turing, A. (1950). Computing machinery and intelligence. ***Mind*, 59**, 433-460. (Dt. Kann eine Maschine denken?, Kursbuch, 8 (1967), 106-138).
- VanLehn, K. (1983). On the representation of procedures in repair theory. In M.P. Ginsburg (Ed.), ***The development of mathematical thinking*** (pp. 197-252). New York, NY: Academic Press.
- VanLehn, K. (1989). ***Mind bugs***. Cambridge, MA: MIT Press.

- VanLehn, K. (1991). Two pseudo-students: Applications of machine learning to formative evaluation. In R. Lewis & S. Otoriki (Eds.), **Advanced research on computers in education** (p. 17-26). Elsevier: North Holland.
- VanLehn, K., Brown, J.S. & Greeno, J. (1984). Competitive argumentation in computational theories of cognition. In W. Kintsch, J. R. Miller & P. G. Polson (Eds.), **Methods and tactics in cognitive science** (pp.235-262). Hillsdale, NJ: Erlbaum.
- VanLehn, K., Jones, R. M. & Chi, M.T. H. (1992). A model of the self-explanation effect. **The Journal of the Learning Sciences, 2**, 1-60.
- Wenger, E. (1987). **Artificial intelligence and tutoring systems**. Los Altos, CA: Morgan Kaufmann.
- Winograd, T. (1972). Understanding natural language. **Cognitive Psychology, 3**, 1-191.
- Winograd, T. (1983). **Language as a cognitive process. Volume I: Syntax**. Reading, MA: Addison-Wesley.
- Winograd, T. & Flores, F. (1986). **Understanding computers and cognition**. Norwood, NJ: Ablex Publishing Corporation. (Dt. **Erkenntnis - Maschinen - Verstehen**. Berlin: Rotbuch, 1989).