

Gedanken zu einer neuen Theorie der Kognition

Rezension von:

Erkenntnis Maschinen Verstehen

T. Winograd / F. Flores:

Michael Mohnhaupt, Klaus Rehkämper

FBI-HH-M-179/90

Mai 1990

Universität Hamburg
Fachbereich Informatik
Bodenstedtstr. 16
D-2000 Hamburg 50

Zusammenfassung

Dieser Beitrag beschäftigt sich mit einer neuen Theorie der Kognition, welche in dem Buch: *Erkenntnis Maschinen Verstehen* von Terry Winograd und Fernando Flores vorgestellt wird. Wir diskutieren einerseits wichtige philosophische Grundannahmen, die der vorgelegten Theorie zugrundeliegen und andererseits wissenschaftstheoretische Aspekte anhand derer wir verschiedene Theorien über Kognition bewerten wollen. Dabei vergleichen wir die Winograd/Flores Theorie mit der alternativen kognitions-wissenschaftlichen Theorie.

Dieser Beitrag erscheint ebenfalls in "Kognitionswissenschaft", 1, Springer-Verlag 1990.

Abstract

This paper is concerned with a new theory on cognition. The theory was introduced by Terry Winograd and Fernando Flores in: *Erkenntnis Maschinen Verstehen*. First, we discuss the philosophical assumptions on which the theory is based. Second, we compare the theory of Winograd and Flores with the 'cognitive science theory' on cognition. The comparison is based on general criteria to evaluate competing scientific theories.

This paper also appears in "Kognitionswissenschaft", 1, Springer-Verlag 1990.

Gedanken zu einer neuen Theorie der Kognition

Michael Mohnhaupt, Klaus Rehkämper,
Universität Hamburg, Fachbereich Informatik,
Bodenstedtstr. 16, D-2000 Hamburg 50

Rezension von: Terry Winograd/ Fernando Flores: *Erkenntnis Maschinen Verstehen. Zur Neugestaltung von Computersystemen*. Übersetzt von Ludwig Voet, mit einem Nachwort von Wolfgang Coy. Berlin: Rotbuch Verlag 1989, 350 S. 42,- DM.

0. Einleitung

In diesem Beitrag beschäftigen wir uns vornehmlich mit der kognitiven Theorie von Terry Winograd und Fernando Flores (im folgenden Winograd/Flores), die sie in ihrem Buch *Erkenntnis Maschinen Verstehen* vorstellen.¹ Dabei werden wir die Gelegenheit benutzen, Kriterien anzugeben, die auch für die Bewertung anderer kognitiver Theorien Gültigkeit besitzen. Diese Kriterien stellen ebenfalls die Grundlage für einen Vergleich mit der konkurrierenden kognitionswissenschaftlichen Theorie dar.

Die Ausführungen von Winograd/Flores berühren gleich vier Gebiete, die zur Entwicklung kognitiver Theorien beitragen können – die Philosophie, die Psychologie, die Neurobiologie und die Künstliche-Intelligenz-Forschung (KI) –, und obwohl Winograd/Flores nicht aus dem Bereich der Philosophie kommen, greifen sie auf philosophische

¹ Dieses Buch ist die deutsche Ausgabe von Winograd, Terry; Flores, Fernando: *Understanding Computers and Cognition: A New Foundation for Design*, Norwood, N. J.: Ablex Publishing Corporation 1986

Traditionen zurück, um auf prinzipielle Probleme innerhalb der KI aufmerksam zu machen. Leider sind diese Traditionen innerhalb der Philosophie nicht unumstritten, und die beiden Autoren erleichtern ihren Lesern und potentiellen Kritikern die Arbeit nicht; die philosophische Diktion, die sie übernehmen, ist ein wenig gewöhnungsbedürftig.

Bevor wir uns jedoch (in Abschnitt 3) mit dem Buch kritisch auseinandersetzen, möchten wir seine Grundgedanken zunächst kurz zusammenfassen (Abschnitt 1) und die Kriterien skizzieren, anhand derer wir Theorien über Kognition bewerten wollen (Abschnitt 2). Dabei werden wir uns auch mit den philosophischen Grundlagen auseinandersetzen. Abschnitt 4 dient der Zusammenfassung der wesentlichen Punkte unserer Kritik an der kognitiven Theorie von Winograd/Flores, und in Abschnitt 5 machen wir einige Anmerkungen zum Nachwort und zur deutschen Übersetzung.

1. *Menschen Erkenntnis Maschinen* - Eine kurze Zusammenfassung

Das Buch gliedert sich in drei Hauptabschnitte: (I) Theoretischer Hintergrund (Kapitel 1-6), (II) Berechenbarkeit, Denken und Sprache (Kapitel 7-10), (III) Gestaltung (Kapitel 11-12). (Alle Verweise und Zitate beziehen sich, wenn nicht anders angegeben, auf die deutsche Ausgabe.) In (I) werden die theoretischen Grundlagen für die späteren Ausführungen bereitgestellt. Gleichzeitig setzen sich die Autoren an dieser Stelle mit der "rationalistischen" Tradition in der Wissenschaft auseinander. In dieser Tradition, die sich bis zu Platon zurückverfolgen läßt, sehen sie die Hauptursache dafür, daß in der bisherigen KI so wichtige Faktoren wie etwa die kontextuelle Einbettung von Sprache keine genügende Berücksichtigung fand. Für diese Kritik können sie

jedoch keine Urheberrechte anmelden, da schon Dreyfus (1979) Platon fälschlicherweise als Ursprung allen Übels identifiziert hat.²

Daß ihre Untersuchung kein leichtes Unterfangen darstellt, ist Winograd/Flores bewußt, und das Spektrum ihres Rasonnierens ist sehr breit angelegt. Sie werden "unweigerlich in Untersuchungen über menschliche Grundphänomene wie 'Intelligenz', 'Sprache' und 'Rationalität' hineingezogen" (24), denn "die Untersuchung, was Computer leisten können, verwickelt uns in Fragen, was Menschen mit Computern anstellen, und schließlich in die grundlegende Frage, was es heißt, ein Mensch zu sein." (25)

Bei ihrer Kritik der rationalistischen Tradition konzentrieren sich Winograd/Flores auf das Phänomen der Sprache. Sprache spielt im Leben des Menschen – des ζῶον λόγον ἔχον – eine zentrale Rolle. "*Nichts existiert außer durch Sprache*" (119). Dieser Rolle wird jedoch, so Winograd/Flores, eine rationalistische, d.h. analytische Beschreibung nicht gerecht, denn einer solchen Beschreibung zufolge ist Sprache nichts anderes als ein Symbolsystem, dessen Symbole (zusammengesetzte Muster) für Dinge dieser Welt stehen, deren Sätze wahr oder falsch, klar oder unverständlich sein können. Das Fundament hierzu bildet die Korrespondenztheorie der Wahrheit, also die Annahme, daß Symbole den Sachverhalten, die sie repräsentieren, entsprechen (40/41).

Winograd/Flores demgegenüber vertreten die Ansicht, daß Sprache nicht der Übermittlung von Information dient (132). Sie akzeptieren im Rückgriff auf Heidegger und Gadamer den hermeneutischen Zirkel: "was wir verstehen, gründet sich auf das, was wir bereits wissen, und was wir wissen, verdanken wir unserer Fähigkeit zu verstehen." (59) Hierbei lassen sich die impliziten Überzeugungen und Anschauungen, die diesem Vorgehen zugrundeliegen, nicht explizit machen (62). Menschen verfügen

² Ganz im Gegenteil war Platon der Ansicht, daß sich nicht das ganze menschliche Wissen in Sprache ausdrücken läßt. Wissen ist nichts, das man wie einen Gegenstand besitzen und einem anderen übergeben kann (vgl. Wieland 1982).

nicht über interne Repräsentationen, die sie mit den tatsächlichen Gegebenheiten vergleichen können. Das Bild, das der rationalistischen Tradition zugrundeliegt – hier der Mensch (Subjekt), dort die Welt (Objekt) – ist ein falsches Bild. "Voreingenommenheit ist kein Zustand, der ein Subjekt die Welt falsch interpretieren läßt, sondern notwendige Voraussetzung dafür, daß überhaupt ein Hintergrund für Interpretation (folglich Sein) vorhanden ist." (61/62)

Menschen sind Teil der Welt, und gerade dieses In-der-Welt-sein bestimmt ihr Wesen. Dadurch daß sie in die Welt 'geworfen' sind, sind sie gezwungen zu handeln, wobei sie die Konsequenzen dieser Handlungen nicht im voraus abschätzen können (66). "Unser In-der-Welt-Sein [ist] kein objektives Spiegelbild der Außenwelt als Vorhandenheit, sondern existiert als Zuhandenheit der Welt, wie sie sich in unseren Handlungen offenbart." (124) Erst wenn diese Zuhandenheit nicht mehr besteht, sie 'zusammenbricht', werden uns die Gegenstände bewußt. So ist z.B. ein Hammer nicht die Menge seiner Eigenschaften, die sich in einer wie auch immer gearteten Repräsentation oder einem mentalen Modell zusammenfassen lassen, sondern er ist die Menge seiner funktionalen Möglichkeiten. Nicht die Eigenschaften sind wichtig, sondern die möglichen Handlungen.

Auch wenn Winograd/Flores "Erkenntnis als biologisches Phänomen" (Kapitel 4) untersuchen, steht das Eingebundensein des Menschen in die Welt im Mittelpunkt der Argumentation. Hierbei übernehmen sie die zentralen Begriffe 'Autopoiesis' und 'strukturelle Kopplung' der Neurobiologen Humberto Maturana und Francisco Varela. Und auch hier wird, nun im Rückgriff auf die Neurobiologie, die Annahme einer Außenwelt und ihrer Repräsentation durch einen objektiven Beobachter verneint. Das Nervensystem als autopoietisches (sich selbst genügendes) Ganzes ist ein nach außen abgeschlossenes System, das keine Eingänge oder Ausgänge besitzt; es kann lediglich durch strukturelle Änderungen dazu veranlaßt werden, neue Zustände einzunehmen. Hierbei ist die Wechselwirkung jedoch nicht dergestalt, daß äußere Zustände zu entsprechenden inneren Zuständen führen, die Außenwelt steuert keine wie auch immer geartete Innenwelt. Veränderungen in der Außenwelt "perturbieren" lediglich das Nervensystem, d.h. sie beeinflussen durch strukturelle

Kopplung die Änderung seiner Struktur. Dabei entwickelt sich ein autopoietisches System notwendig so, daß seine Aktivitäten zweckmäßig an sein eigenes Umfeld angekoppelt sind (82), und "Lernen ist kein Prozeß der Akkumulation von Repräsentationen der Umwelt." (Maturana 1982, zit. nach Winograd/Flores: 82)

Der Begriff *Sprache* wird von Maturana sehr weit gefaßt und umspannt nahezu alle möglichen Arten der Interaktion. Sie existiert in einer Gemeinschaft von Individuen und erneuert sich ständig durch den Gebrauch und die damit verbundene strukturelle Kopplung (89). Nach Maturanas Überzeugung ist Sprache ein definierendes Merkmal menschlichen Daseins, denn "Sprache (gemeinschaftlich, nicht privat) ist ausschlaggebend für bewußtes Denken." (93)

Sprechen ist eine zentrale Form von Handeln, und "Menschsein bedeutet, einer Gattung anzugehören, die durch Sprechen und Zuhören wechselseitige Verpflichtung herstellen kann. ... Computer sind jedoch ungeeignet, Verpflichtungen einzugehen und können sich daher nicht selbst am Sprachprozeß beteiligen." (132f.)

Auch der dritte Bereich, den Winograd/Flores für ihre Untersuchung fruchtbar machen wollen, beschäftigt sich mit Sprache und sozialer Interaktion. Die Sprechakttheorie Austins und Searles trägt der Tatsache Rechnung, daß Sprechen Handeln ist und man durch Sprache Verpflichtungen eingeht. (Searle 1969) Aber, obwohl die Sprechakttheorie verschiedene Sprechhandlungen zu unterscheiden weiß und die verschiedenen Rollen von Sprechern und Hörern berücksichtigt, ist sie in der rationalistischen Tradition verhaftet und geht nicht über sie hinaus. Ein grundlegender Fehler ist, nach Winograd/Flores, z.B. die Ansicht, daß Sätze unabhängig vom Kontext eine wörtliche Bedeutung besitzen. Dennoch verwenden Winograd/Flores eine Menge Zeit und Papier, diese Theorie vorzustellen und Beispiele anzuführen, die das Gegenteil beweisen sollen. So gestehen sie zwar zu, daß Searle die Bedeutung des Hintergrundwissens erkannt hat, sehen aber in seiner Formulierung des Problems eine unzulässige Verharmlosung. Man könnte jedoch, so ihre Befürchtung, den Eindruck gewinnen, daß der Hintergrundbezug durch

eine formal-semantische Beschreibung erfaßt werden kann. Dies träge jedoch nicht den Kern, denn er ist "keine Sammlung von Bedeutungssätzen (Propositionen), sondern unsere Grundorientierung von 'Sorge'." (102) Der Grundaussage der Sprechakttheorie, daß Sprechen das Eingehen von Verpflichtungen beinhaltet, stimmen sie jedoch zu. Sie schließen sich der Auffassung Habermas' an, daß "jede Sprachhandlung für die Beteiligten Konsequenzen zeitigt. Eine Sprachhandlung initiiert weitere, unmittelbare Handlungen und erzeugt Bindungen für zukünftige Handlungen." (105)

Hiermit sind die Grundlagen "für eine neue Orientierung gelegt" (Kapitel 6). Die zentralen Begriffe des 'In-der-Welt-sein' und der 'Zuhandenheit' (Heidegger) bzw. der 'strukturellen Kopplung' (Maturana) sind eingeführt. Die einfältige Annahme einer objektiven Außenwelt (126), die mittels interner Repräsentationen erfaßt werden kann, wurde verworfen. Gleichzeitig wurde die Rolle des Hintergrundbezuges beim Verstehen deutlich gemacht. Sprache und Erkenntnis sind keine mentalen Phänomene, sondern soziale Handlungen. Wesentliches Merkmal menschlicher Existenz ist die Fähigkeit, Verpflichtungen einzugehen, bzw. eingehen zu müssen.

Dieses zentrale Merkmal greifen Winograd/Flores im zweiten Teil des Buches "Berechenbarkeit, Denken und Sprache" auf und wenden es auf die Idee einer künstlichen Intelligenz an. Hierbei gelangen sie zu dem Schluß, daß Computer die Fähigkeit fehlt, Verpflichtungen eingehen zu können und sie sich somit auf einer völlig anderen Seinsebene befinden (179).

Reflexion ist ohne Repräsentation nicht möglich. Schon die Strukturierung der Umwelt mit Hilfe von Begriffen stellt eine solche Repräsentation dar. Menschen können und müssen jedoch in ihrer Umwelt ohne die Möglichkeit der Reflexion handeln. Sie bedienen sich zwar manchmal solcher Repräsentationen, diese sind jedoch keine unabdingbare Voraussetzung (167). Reflexion und auch Abstraktion sind zwar wichtige Phänomene unseres Daseins, aber sie sind auch verantwortlich für eine gewisse, perspektivische 'Blindheit'. "Unsere

Sicht ist eingeschränkt auf das, was wir mit diesen gegenständlichen Begriffen ausdrücken können." (164)

Computer hingegen sind ohne Repräsentationen nicht denkbar. Daher ist das Phänomen der 'Blindheit' in diesem Bereich unumgänglich. Winograd/Flores bemühen sich nun, in den einzelnen Kapiteln "Computer und Repräsentation" (7), "Berechenbarkeit und Intelligenz" (8), "Sprache verstehen" (9) und "Aktuelle Entwicklungstendenzen in der Künstlichen Intelligenz" (10) diese Problematik detailliert aufzuzeigen. Hierbei werden die Bemühungen, durch die Verwendung nicht-monotoner Logiken Fortschritte zu erzielen, genauso negativ beurteilt wie Winograds eigenes Programm SHRDLU. Das Nachvollziehen und Beurteilen dieser Argumentation möchten wir jedoch dem Leser überlassen.

Quintessenz der Ausführungen ist, daß mechanische Geräte, z.B. Wecker, und Computer keine prinzipiellen Unterschiede aufweisen (158). Menschen zeichnen sich jedoch dadurch aus, daß sie verstehen können. Dies bedeutet allerdings nicht, daß sie Repräsentationen manipulieren, denn "Verstehen ist keine festgefügte Beziehung zwischen Repräsentation und repräsentierten Gegenständen, vielmehr die Verpflichtung zur Teilnahme an einem Dialog, der vor dem übergreifenden Horizont von Sprecher und Hörer stattfindet, in einer Weise, die neue Unterscheidungsmerkmale entstehen läßt (207). Nicht zu repräsentieren, sondern Verpflichtungen einzugehen, ist das Wesentliche der Sprache, "aber wie soll ein Computer Bindungen eingehen?" (205)

Für die gegenwärtige KI ergeben sich aus diesen Ansichten keine optimistischen Prognosen, die grandiosen Ziele sind unerreichbar. Allerdings werden nützliche Nebenprodukte abfallen. Diesen widmen sich Winograd/Flores im letzten Teil des Buches mit dem Titel "Gestaltung".

Hier setzen sich Winograd/Flores zunächst (Kap. 11) damit auseinander, welchen Nutzen Computersysteme für Entscheidungsträger im Management mit sich bringen können, und im letzten Kapitel (Kap. 12) leiten sie Auswirkungen ihrer Theorie auf den Entwurf von Computersystemen ab. Es werden einige wichtige Anmerkungen gemacht. So wird zum

vorsichtigen Umgang mit dem Übertragen von Entscheidungen auf Computer gemahnt, und es wird immer wieder auf die große Bedeutung des Kontextes und der sozialen Einbettung von Aktionen und Handlungen hingewiesen. Beides folgt aus den Erkenntnissen der ersten beiden Teile des Buches, aber es wäre ein logischer Fehler, von der Gültigkeit der Einsichten im Kapitel "Gestaltung" auf die Gültigkeit der kognitiven Theorie von Winograd/Flores zu schließen. Wir gehen im Folgenden nicht weiter auf das Kapitel "Gestaltung" ein, weil wir uns hauptsächlich mit der kognitiven Theorie auseinandersetzen wollen.

2. Bewertungskriterien

In diesem Abschnitt geben wir die Kriterien an, nach denen wir Theorien über Kognition bewerten wollen. Wir teilen diese Kriterien in zwei Teilaspekte 'philosophische Grundlagen' (2.1) und 'Theorienbewertung' (2.2) auf.

2.1 Philosophische Grundlagen

Es ist nicht ganz einfach, die Grundannahmen, die Winograd/Flores sich von verschiedenen Autoren aus unterschiedlichen Bereichen ausgeborgt haben, hier in aller Ausführlichkeit zu diskutieren. Zumal die einzelnen Standpunkte auch in ihren jeweiligen Disziplinen nicht unumstritten sind. Philosophische Gedankengebäude lassen sich sicherlich nicht so einfach klassifizieren und bewerten, wie dies bei wissenschaftlichen Theorien der Fall ist; es gibt keine allgemein akzeptierten metatheoretischen Kriterien. Somit sind die Ausführungen von Heidegger, Gadamer, Habermas und Searle nicht einfach in richtig oder falsch, zutreffend oder verfehlt einzuteilen. Dennoch kann man sich fragen, inwieweit die vorgetragenen Systeme in der Lage sind, die Grundbegriffe einer Theorie, in diesem Fall der Theorie von Winograd/Flores (s.u. 2.2), plausibel werden zu lassen. Hierbei gehen wir davon aus, daß es eine der vornehmsten

Aufgaben der Philosophie ist, die Grundlagen und -begriffe der einzelnen Wissenschaften zu überprüfen, zu hinterfragen und in einen sinnvollen Zusammenhang zu bringen. Philosophie ist keine Wissenschaft, aber Wissenschaft ohne Philosophie ist nicht möglich.

Etwas anders verhält es sich bei Maturana und seinen Gedanken. Maturana ist Vertreter einer wissenschaftlichen Disziplin, und auf ihn und seine Äußerungen lassen sich die Kriterien anwenden, die wir in 2.2 angeben.

Bei allem, was nun folgt, sind wir uns im klaren darüber, daß die Autoren uns vorhalten würden, daß wir immer noch tief in die rationalistische Tradition verstrickt sind und daher die von ihnen vorgetragenen Gedanken überhaupt noch nicht verstanden haben (Vgl. ihre Antwort auf verschiedene Rezensionen in 'Artificial Intelligence' Vol. 31, 1987, Seite 213-261). Allerdings sind wir davon überzeugt, daß diese Art der Immunisierung gegenüber abweichenden Ansichten einer wissenschaftlichen Diskussion nicht förderlich ist.

Wir teilen mit Winograd/Flores die Auffassung, daß die bisherige KI die philosophischen Grundlagen zu wenig berücksichtigt hat. Daher sind wir der prinzipiellen Fragestellung des Buches sehr zugetan. Nur die hier präsentierten "Antworten" möchten wir nicht übernehmen.

2.2 Theorienbewertung

Der umfangreichste und entscheidende Beitrag des Buches (Teil I und II) besteht aus einer Theorie über menschliche Kognition. Winograd/Flores versuchen u.a. zu erklären, wie Menschen Wahrnehmungen und Sprache verstehen, und wie sie auf veränderliche Situationen in der Welt reagieren und Entscheidungen fällen. Dabei ist es den Autoren wichtig, in ihrer Theorie auf metaphysische Elemente zu verzichten, sie bemühen sich also um Modellbildung innerhalb der empirischen Wissenschaften.

"Wir wollen keiner Verteidigung der Irrationalität oder der esoterischen Anziehungskraft nichtrationaler Eingebungen Vorschub leisten. ... Wir

fühlen uns verpflichtet, eine neue Grundlage für Rationalität zu entwickeln – eine Grundlage, die in ihrer Zielsetzung genauso präzise ist wie die rationalistische Tradition, die aber deren zugrundeliegenden Voraussetzungen nicht teilt." (27)

Die Theorie von Winograd/Flores tritt damit in direkte Konkurrenz zu anderen kognitiven Theorien, insbesondere und beabsichtigt in Konkurrenz zum Forschungsprogramm der Kognitionswissenschaften:

"In unserer Kritik ist damit auch implizit eine Stellungnahme zur Forschungseinrichtung der Kognitionswissenschaften enthalten. Wir halten diese zwar nicht für geistlos und unnützlich, aber ihr wird es doch wesentlich an Weitblick und Fähigkeiten fehlen, um zur Klärung der Frage, was wir sind und was wir tun, beitragen zu können." (54)

Die wesentlichen Hypothesen der Kognitionswissenschaften werden verworfen, entweder durch Aussagen von Winograd/Flores selbst, oder von Autoren, auf welche sich Winograd/Flores berufen. Dabei handelt es sich hauptsächlich um Maturana, Varela, Heidegger und Gadamer. Insbesondere werden die zwei Grundhypothesen der Kognitionswissenschaften abgelehnt:

1. Die Hypothese, daß kognitive Prozesse auf Informationsverarbeitung beruhen.

"Er (Maturana) verwirft die Metapher von der Informationsverarbeitung als Grundlage für Erkenntnis und ersetzt die Frage "Wie gelangt der Organismus an die Information über seine Umwelt" durch die Frage "Wie kommt es, daß der Organismus über eine Struktur verfügt, die es ihm erlaubt, im Milieu, das ihn umgibt, angemessen zu wirken". (84)

"Eine wichtige Konsequenz der Kritik im ersten Teil des vorliegenden Buches ist, daß Sprache keineswegs als bloße Übermittlung von

Information angesehen werden kann." (132), [im Original: "...language cannot be understood as the transmission of information" (76)]³

2. Die Hypothese, daß mentale Repräsentationen wichtiger Teile der Außenwelt⁴ ein zentraler Bestandteil kognitiver Prozesse sind.

"Wir beziehen uns nicht in erster Linie über Vorstellungen auf Dinge" (63), [im Original: "We do not relate to things primarily through having representations of them. ... Heideggers rejection of mental representations." (33)]

"Die Strukturen des Nervensystems repräsentieren seiner (Maturanas) Auffassung nach keineswegs die Welt, in der der Organismus lebt." (156)

Für eine Auseinandersetzung mit der vorliegenden Theorie bietet es sich daher an: 1) die Leistungsfähigkeit der Theorie an anerkannten Kriterien zur Theorienbewertung zu messen (siehe z.B. Stegmüller 1979), und 2) die einzelnen Hypothesen der Theorie mit entsprechenden Hypothesen der konkurrierenden kognitionswissenschaftlichen Theorie zu vergleichen. Insofern ist unsere Argumentation sowohl eine Auseinandersetzung mit der kognitiven Theorie von Winograd/Flores, als auch mit wichtigen Aspekten der kognitionswissenschaftlichen Theorie, welche wir für die zur Zeit tragfähigste kognitive Theorie halten.

Stegmüller (1979) integriert verschiedene Ansichten (im Wesentlichen die von Popper (1936), Kuhn (1970) und Sneed (1971)) zur Evolution von Theorien und kommt zu dem Schluß, daß man Theorien sinnvoll in die zwei Bereiche einteilen sollte: *Strukturkern* und *empirische Komponenten*. Während die empirischen Komponenten (die Menge der Anwendungen der Theorie) widerlegbar sind, entzieht sich der

³ Die Probleme der Übersetzung behandeln wir noch einmal gesondert in Kap. 5.2.

⁴ Die Existenz einer Außenwelt, die zumindest teilweise zugänglich ist, wird von uns konzediert, obwohl wir uns darüber im klaren sind, daß diese Position (hypothetischer Realismus) nicht empirisch beweisbar ist.

Strukturkern (die Fundamentalgesetze) einer empirischen Falsifikation. Eine neue Theorie kann eine alte Theorie nur dann erfolgreich (mit Erkenntnisfortschritt) *verdrängen*, wenn ihre Erklärungen und Voraussagen diejenigen der alten Theorie beinhalten und darüber hinausgehen.

Bei der Bewertung der unterschiedlichen empirischen Komponenten ist es wichtig, auf folgende Kriterien zu achten (siehe auch Hempel 1966, Carnap 1966):

1. *Falsifizierbarkeit* ist eine notwendige Anforderung an empirische Hypothesen (nicht auf den strukturellen Kern der Theorie).
2. *Empirische Tragweite*. Es ist zu prüfen, ob die vorhandenen empirischen Ergebnisse mit den Hypothesen im Einklang stehen, ihnen widersprechen, oder in keinem Zusammenhang mit ihnen stehen.
3. Die *Qualität* und *Vielfalt* der empirischen Stützung.
4. Außerdem ist die *Einfachheit* von empirischen Hypothesen ein wichtiges Vergleichskriterium. Einfachheit geht in vielen Fällen mit leichter Falsifizierbarkeit einher (Popper 1936).
5. Die *Vorhersagekraft* einer empirischen Hypothese bezieht sich auf die Möglichkeit, Aussagen über noch unbekannte Tatsachen zu machen, und damit verbundene empirische Tests zur Bestätigung oder Widerlegung anzugeben.

Ein Vergleich konkurrierender empirischer Komponenten von Theorien anhand der genannten 5 Punkte trägt entscheidend zur Einschätzung bei, ob eine Theorienverdrängung sinnvoll ist.

Wir hoffen aus zwei Gründen, daß die hier beschriebenen Kriterien zur Theorienbewertung auch von Winograd/Flores im Prinzip (eventuell mit anderer Gewichtung) akzeptiert werden können. Erstens, weil Winograd/Flores teilweise bemüht sind, ihr Modell mit beobachtbaren Alltagsphänomenen empirisch zu begründen, bzw. weil sie ihr Modell im

dritten Teil ihres Buches auf praktische Probleme anwenden. Und zweitens, weil sie metaphysische Erklärungen explizit ablehnen.

3. Kritik der Theorie von Winograd/Flores

3.1 Bewertung der philosophischen Grundlagen

Wir müssen gestehen, daß die Ausführungen Heideggers für uns nicht immer verständlich sind. Dies gilt insbesondere im Hinblick auf die Frage, was sich daraus für eine wissenschaftliche Theorie der Kognition ergibt. Als Ausdruck eines Lebensgefühls, als Poesie mögen sie von großem Wert sein; für die unter 2.1 aufgeworfenen Fragen bleibt uns dieser Wert jedoch verborgen. Wir schließen uns der Auffassung Platons an, die er z.B. im *Protagoras*, in der *Apologie*, aber auch im *Ion* geäußert hat, nämlich, daß Dichtkunst und Wissenschaft sich deutlich unterscheiden und klar voneinander getrennt bleiben sollten (vgl. Wieland 1982, Bröker 1967²). Dichter sprechen über die Dinge und können sie doch nicht begründen; sie können keine diesbezüglichen Fragen beantworten und auch einem Fragenden nicht weiterhelfen. Die Quelle ihrer Einsichten bleibt unklar.

Heideggers Ansichten sind schon früh kritisiert worden. Hierbei kam die Kritik aus einer Richtung, die Winograd/Flores sicherlich zur rationalistischen Tradition rechnen würden. So beschäftigt sich Rudolf Carnap in seiner Arbeit *Zur Überwindung der Metaphysik durch logische Analyse der Sprache* ausführlich mit den Gedanken Heideggers. Und man muß, obwohl der Ansatz des Wiener Kreises im Laufe von nunmehr fast sechzig Jahren einige berechtigte Wandlungen durchlaufen hat, der Grundposition der Carnapschen Kritik beipflichten: Die Wörter eines Satzes müssen eine Bedeutung haben und bei Sätzen muß man angeben können, wann sie wahr oder falsch sind. Sicherlich ist die Ansicht, daß die Bedeutung eines Satzes in seiner Verifikation liegt, nicht haltbar, aber dies ist auch nicht die Ansicht Wittgensteins, auf den sich Carnap beruft, denn im *Tractatus* (4.024) heißt es: "Einen Satz verstehen, heißt, wissen

was der Fall ist, wenn er wahr ist. (Man kann ihn also verstehen, ohne zu wissen, daß er wahr ist.) Man versteht ihn, wenn man seine Bestandteile versteht." Wir müssen aber gestehen, daß uns der Sinn eines Satzes der folgenden Art verborgen bleibt:

Der Satz: "Der Mensch existiert" bedeutet: Der Mensch ist dasjenige Seiende, dessen Sein durch das offenstehende Innestehen in der Unverborgenheit des Seins, vom Sein her, im Sein ausgezeichnet ist. (Heidegger 1929: 16)⁵

Dies ist zwar kein Zitat, das Winograd/Flores verwenden, aber wir gehen davon aus, daß Winograd/Flores den gesamten Heideggerschen Ausführungen positiv gegenüberstehen, und gleichzeitig ist dieses Zitat typisch für die von Winograd/Flores adoptierten Heideggerschen Gedanken. Jedoch glauben auch wir, wie Wolfgang Coy es im Nachwort sehr plastisch ausdrückt, daß den beiden Autoren eine gewisse "Steinbruch-Mentalität" (309) zueigen ist. Dadurch ist es nicht unbedingt sichergestellt, daß die Teile, die übernommen werden, eine solche Amputation bzw. Verpflanzung unbeschadet überstehen. So gehen Winograd/Flores in ihren Ausführungen nicht auf zwei der zentralen Heideggerschen Begriffe – die *Angst* und das *Nichts* – ein. Aber gerade den Begriff des Nichts bzw. die Möglichkeit ihn als logisches Subjekt einer Aussage verwenden zu können hat Carnap in der oben schon erwähnten Arbeit zu analysieren versucht, allerdings ohne Erfolg. Auch das angeführte Beispiel widersteht einer solchen Analyse; obwohl man den Ausdrücken des Satzes, läßt man das 'principle of charity' walten, jeweils einen Sinn zugestehen kann, gilt dies nicht für den ganzen Satz. Es erscheint uns unmöglich, Wahrheitsbedingungen für diesen Satz anzugeben.

⁵ Wir geben zu, daß sich eine annähernde prädikatenlogische Übertragung des Satzes "Der Mensch existiert" in $[(\exists x)(x \text{ ist ein Mensch})]$ bzw. $(\exists x)(M(x))$, mit $M = \text{menschlich}$ sehr profan neben der Heideggerschen Wortgewalt ausmacht.

Wie steht es nun mit einer anderen zentralen These der Heidegger-Rezeption durch Winograd/Flores; der Tatsache, daß Gegenstände nicht als ein Bündel von Eigenschaften verstanden werden sollen, sondern daß ihre Funktionalität im Vordergrund steht. "Meine Handlungsfähigkeit ist Resultat der Vertrautheit mit dem *Hämmern*, nicht meines Wissens über einen *Hammer*." (64, Hervorhebung im Original)

Diese Ansicht wird auch innerhalb der Entwicklungspsychologie vertreten. Die Ausbildung von drei verschiedenen Repräsentationsformen läßt sich innerhalb der kindlichen Entwicklung nachweisen (vgl. z.B. Aebli 1980, 1981). Am Anfang steht die aktionale oder handlungsmäßige Repräsentation, Bruner nennt dies auch die enaktive Repräsentation; ihr folgt die ikonische oder bildhafte Repräsentation, der sich die symbolische (sprachliche) Repräsentation anschließt. In ihr treten die figurativen Elemente, die bei der bildhaften Repräsentation im Vordergrund standen, wieder zugunsten der funktionalen Aspekte eines Gegenstandes zurück. Diese Form der Repräsentation knüpft also wieder an die aktionale Repräsentation an, bei der die sensumotorischen Eigenschaften der Objekte ausschlaggebend waren. Man kann also feststellen, daß diese Beobachtungen Heideggers in der Wissenschaft nicht unbeachtet blieben. Allerdings sind sie in eine repräsentationale Theorie des Wissens, die deklarative und prozedurale Aspekte enthält, integriert worden und bieten somit keinen prinzipiellen Einwand gegen die Möglichkeit, Wissen mittels mentaler Repräsentationen zu verarbeiten.

Der Verweis auf den hermeneutischen Zirkel ist sicherlich nicht unberechtigt, aber die Frage bleibt, wie berechtigt er an dieser Stelle ist. Niemand wird bestreiten, daß wir mit einem gewissen Vorverständnis und gewissen Erwartungen an einen Text herangehen, und gerade dies ist ja durch Untersuchungen innerhalb der Psychologie und der Psycholinguistik deutlich geworden (vgl. z.B. Bartlett 1932, Sanford & Garrod 1981, Johnson-Laird 1983). Diese Untersuchungen und die darauf aufbauenden Überlegungen haben aber mittlerweile auch Eingang in die KI gefunden (z.B. Schank & Abelson 1977; Eschenbach et al. 1990). Wir sympathisieren daher sehr stark mit der Aussage von André Vallinos, der

in seiner Rezension des Buches von Winograd/Flores einen Ausschnitt, der sich mit dem Problem des Hintergrundbezuges beschäftigt, zitiert und feststellte: "Frankly, I do not understand what this passage says that cannot be stated more simply by the following sentence: 'meaning depends on a set of background beliefs, desires, and dispositions.'" ('Artificial Intelligence' Vol. 31, 1987, Seite 217)

Maturanas Thesen bilden mittlerweile sogar die Grundlage für eine neue philosophische Richtung – die Philosophie des *Radikalen Konstruktivismus* (vgl. u.a. Schmidt 1987). Diese neue Richtung, die sich in ihrem Letztbegründungsanspruch als direkte Konkurrenz zur *Evolutionären Erkenntnistheorie* auffaßt, ist jedoch schon an anderer Stelle erfolgreich kritisiert worden (Wendel 1989). Wendel zeigt, daß sich die These, daß der Mensch ein autopoietisches System sei, nicht konsistent aufrecht erhalten läßt. Die Annahmen, die dem Radikalen Konstruktivismus zugrundeliegen, sind zirkulär. Der Radikale Konstruktivismus gelangt durch empirische Untersuchungen zu der Einsicht, daß der Mensch ein autopoietisches System ist und leitet davon die Erkenntnis ab, daß eine 'wirkliche' Realität nicht existiert. Dies führt allerdings zu dem Schluß, daß die vorgenannten, wissenschaftlichen Untersuchungen nichts über die Wirklichkeit aussagen. Wendel schreibt daher: "Wir sehen, daß die radikal konstruktivistische Erkenntniskonzeption nur solange eine gewisse Plausibilität für sich in Anspruch nehmen kann, als sie die eigenen Voraussetzungen unter der Hand realistisch deutet, was sie jedoch inkonsistent machen würde." (1989: 88)

Ein letzter Punkt, den wir an dieser Stelle ansprechen möchten, betrifft die These, daß Computer prinzipiell keine Verpflichtung eingehen können. Verpflichtungen eingehen zu können, ist das Vorrecht von Wesen, die eine Sprache sprechen und verstehen, eben von Menschen. Hier zeigt sich u.E. deutlich ein Problem in der Fragestellung, die Winograd/Flores bearbeiten. Intelligenz, Verstehen und Menschsein werden scheinbar als identisch angesehen. Winograd/Flores sind nicht in der Lage, Kriterien für Intelligenz oder Verstehen anzugeben, die unabhängig vom Menschen formuliert sind. Diese anthropozentrische Sicht führt jedoch zu der Annahme, daß die KI darauf aus ist, einen

künstlichen Menschen zu bauen. Dies ist aber nicht der Fall. Für uns ist daher die entscheidende Frage: Welche Kriterien haben wir, nicht-menschlichen Objekten (oder Wesen) Intelligenz, die Möglichkeit zu verstehen und Verpflichtungen eingehen zu können, zuzuschreiben.

Vielleicht gewinnt die Frage an Deutlichkeit, wenn wir den Bereich menschlicher Artefakte – z.B. Computer – einmal verlassen. In Science Fiction Romanen ist dieses Szenarium schon tausendmal verwendet worden; außerirdische Wesen, ob gut oder böse lassen wir einmal dahingestellt, landen auf der Erde (vorzugsweise in den Vereinigten Staaten) und nehmen Kontakt zur Erdbevölkerung auf. Hier sind wir auf äußere Kriterien angewiesen, die uns angeben, ob diese Wesen intelligent sind und Verpflichtungen eingehen können. Wir können sie nicht auseinanderschrauben, um zu sehen, ob sie etwa aus Silizium aufgebaut sind oder ob sie das Sprachverstehen vielleicht nur mittels verschiedenartiger Repräsentationen simulieren. Wenn sie nach unserem Verständnis sinnvoll handeln, und zu diesem Handeln zählen wir auch das Sprechen, dann haben wir ausreichend Evidenz für die Annahme, daß es sich um nicht-menschliche, intelligente Wesen handelt. Aufgrund dieses Gedankenexperiments halten wir die Aussage von Winograd/Flores, daß Computer sich auf einer anderen Seinsebene befinden und daher prinzipiell niemals in der Lage sein werden, Verpflichtungen eingehen zu können, nicht für sehr überzeugend. Möglicherweise kann man den beiden Autoren zugestehen, daß sie die Frage untersuchen wollten, was es bedeutet, ein Mensch zu sein; die Frage, was es bedeutet, intelligent zu sein, haben sie jedenfalls noch nicht einmal im Ansatz beantwortet.

3.2 Wissenschaftstheoretische Aspekte und ein Vergleich mit der kognitionswissenschaftlichen Theorie

Wir wollen hier keine erschöpfende Bewertung aller Einzelhypothesen der Theorie von Winograd/Flores durchführen, sondern uns einige wesentliche Beispiele herausgreifen.

Als Schwachpunkte der Theorie von Winograd/Flores sehen wir die geringe empirische Tragweite und die damit verbundene mangelnde oder nur vage Möglichkeit der Falsifikation empirischer Komponenten der Theorie an. Empirische Ergebnisse aus psychologischen Forschungen werden von Winograd/Flores fast überhaupt nicht zur Stützung ihrer Theorie herangezogen. Eine Ausnahme bilden Ergebnisse über die Wahrnehmung von Farben, welche auf Arbeiten von Maturana basieren (Kapitel 4). Psychologische Experimente zeigen, daß keine einfache Zuordnung zwischen emittierter Wellenlänge eines Objekts und wahrgenommener Farbe des Objektes existiert. Daraus schließen Maturana und Varela: 1) daß die Farbwahrnehmung bei Menschen einen hohen konstruktiven (vom System und nicht vom Stimulus bestimmten) Anteil hat, und 2) daß diese konstruktive Interpretation auf die gesamte Kognition verallgemeinerbar ist.

Der erste Schluß ist möglich, aber keineswegs notwendig. Es ist mindestens ebenso plausibel, daß die Farbwahrnehmung ausschließlich vom physikalischen Stimulus abhängt, nur erheblich komplexer, als von lokalen Oberflächen- und Beleuchtungseigenschaften. Aber selbst wenn sich die erste Folgerung von Maturana und Varela als zutreffend erweisen sollte, halten wir den zweiten Schluß, vom Farbsehen auf die gesamte Kognition zu verallgemeinern, für unzulässig. Es ist bekannt, und fast jedem Textbuch über Wahrnehmung zu entnehmen, daß das Farbsehen u.a. wegen der genannten nichtlokalen Eigenschaften eine Sonderrolle im visuellen System einnimmt. Andere Verarbeitungsprozesse auf derselben Ebene, z.B. Orientierungs-, Bewegungs- und Tiefenwahrnehmung hängen nachweisbar lokal und direkt vom physikalischen Stimulus ab, eine konstruktivistische Interpretation läßt sich darauf nicht aufbauen. Darüber hinaus ist die funktionelle Bedeutung des Farbsehens erheblich geringer als die der anderen Prozesse auf der gleichen Verarbeitungsebene. Denn ein Ausfall des Farbsehens kann weitgehend verkraftet werden, ohne Orientierungs- oder Bewegungssehen ist eine Wahrnehmung der Welt drastisch eingeschränkt.

Die empirische Begründung der kognitiven Theorie von Winograd/Flores mithilfe des Farbsehens halten wir deshalb für schwach. Andere empirische Ergebnisse werden gar nicht erst berücksichtigt, obwohl tausende von Resultaten in der kognitiven Psychologie erarbeitet worden sind. Winograd/Flores halten diese Ergebnisse offensichtlich für so reduktionistisch (Kap. 2.1), daß deren Beachtung nicht lohnt. Dennoch müssen diese Ergebnisse von einer umfassenden Theorie erklärbar sein, zumindest als Randphänomene. Leider schlagen Winograd/Flores selber keinerlei Versuche vor, weder zur Bestätigung noch zur Widerlegung ihrer Theorie.

Ob es prinzipiell möglich ist, die empirischen Komponenten der Theorie von Winograd/Flores zu falsifizieren, bleibt auch an folgender Stelle unklar. Wie oben schon erwähnt, diskutieren Winograd/Flores vier unterschiedliche Rezensionen der englischsprachigen Ausgabe des vorliegenden Buches ('Artificial Intelligence' Vol. 31, 1987, Seite 213-261):

André Vellino sieht in der neuen Theorie keine überzeugende Antwort auf die bekanntermaßen schwierige Probleme beim Sprachverstehen. Er resümiert: "Their conclusions promise to be devastating, but they are in fact quite meek". Mark Steffic und Daniel Bobrow halten die Hauptaussagen ebenfalls für schwach, und sehen die Kritik an kognitions-wissenschaftlichen Modellen als sehr unpräzise an. Lucy Suchman lobt das Buch für viele berechtigte Fragen, und William Clancey hält den Beitrag für einen wissenschaftlichen Meilenstein vom Kaliber kopernikanischer oder freudscher Arbeiten.

Winograd/Flores antworten den verschiedenen Kritiken folgendermaßen:

"The reviews collected here offer a striking validation of this theory. The same piece of language (in this case our book) produced four widely different understandings, each generated within the background of a particular listener".

Dieses Zitat beinhaltet eine Folgerung, die uns sehr verwundert. Aus der Tatsache, daß verschiedene Personen den gleichen Text unterschiedlich

verstehen, wird die Bestätigung der von Winograd/Flores vertretenen Theorie gefolgert. Offenbar besteht folgende Alternative: entweder man akzeptiert oder verwirft einstimmig die Gültigkeit der Winograd/Flores Theorie, oder man ist unterschiedlicher Meinung, aber jede Art von Interpretations- oder Meinungsunterschied belegt die Winograd/Flores Theorie!

Wir ziehen hier den Schluß, daß die empirischen Komponenten der kognitiven Theorie von Winograd/Flores gegen Falsifikation weitgehend immunisiert sind.

Auch andere Kriterien zur Theorienbewertung verbessern den Eindruck der Winograd/Flores Theorie nicht. Die Einfachheit der Theorie wirkt zunächst positiv, denn sie besteht aus wenigen Grundkonzepten, wie 'strukturelle Kopplung' 'autopoietisches System', 'Geworfenheit', usw., und die *internen Prinzipien* der Theorie, d.h. die Bezüge der verwendeten Begriffe untereinander, sind einigermaßen nachvollziehbar. Aber die *Brückenprinzipien* der Theorie, d.h. die Bezüge der Begriffe zu konkreten empirischen Erkenntnissen, sind leider nicht erkennbar. Ohne Brückenprinzipien hat eine Theorie weder erklärende Kraft, noch Vorhersagekraft (Hempel 1966). Auch Vielfalt und Qualität der Theorie sind daher wegen fehlender empirischer Tragweite kaum zu beurteilen.

Zwei Zitate sollen verdeutlichen, warum wir Brückenprinzipien der Winograd/Flores Theorie vermissen:

1) "Eher als durch Übermittlung von Information entsteht Bedeutung durch aktives Zuhören" (102) [im Original: "Meaning is created by an active listening, in which the linguistic form triggers interpretation rather than conveying information (57)],

2) "Nicht ein Modell vom Geisteszustand des Sprechers oder Zuhörers wird hier entworfen, sondern die Konversation in Form einer Tanzfigur dargestellt (114).

Es wird nicht deutlich, was es konkret bedeutet, 'Interpretationen zu triggern' oder 'zu tanzen', wie sich diese Phänomene gegen andere

Phänomene abgrenzen, welche empirischen Ergebnisse mit diesen Begriffen erklärt werden können, und wie diese Aussagen eventuell widerlegt werden können.

Die insgesamt harte Kritik an der Winograd/Flores Theorie wirft die Frage auf, ob die kognitionswissenschaftliche Theorie anhand der genannten Kriterien besser abschneidet. Wir können dies hier nicht in aller Tiefe prüfen, werden uns aber einige typische Anwendungen der Grundhypothesen der kognitionswissenschaftlichen Theorie herausgreifen.

Zunächst wenden wir uns der Informationsverarbeitungshypothese zu. Die Gültigkeit dieser Annahme wird für viele Anwendungsbereiche beansprucht, z.B. für visuelle Wahrnehmung, höhere kognitive Prozesse, Erinnerung, etc., und eine sehr große Anzahl empirischer Ergebnisse aus vielen Bereichen der Kognition ist mit dieser Annahme verträglich (siehe z.B. Boff, Kaufman & Thomas 1985). Außerdem resultieren aus der Informationsverarbeitungsannahme falsifizierbare empirischen Vorhersagen. So haben für Informationsverarbeitungsprozesse Komplexitätsbetrachtungen notwendigerweise Gültigkeit. Anhand von zwei Beispiele skizzieren wir empirische Aussagen, welche aus Komplexitätsüberlegungen resultieren:

1) Wenn die Wahrnehmung topologischer Eigenschaften zeitlich eher möglich ist als die Wahrnehmung elementarer Formeigenschaften, muß die Informationsverarbeitungshypothese bezweifelt werden (siehe auch Mohnhaupt & Neumann (1990)). Denn aus Informationsverarbeitungssicht ist die Wahrnehmung topologischer Eigenschaften erheblich komplexer als die Wahrnehmung einfacher Formmerkmale. Chen (1982, 1989) präsentiert Untersuchungen, die eine schnellere Wahrnehmung topologischer Eigenschaften experimentell zu belegen scheinen. Im Rahmen einer Theorie, welche mit der Informationsverarbeitungshypothese verträglich ist, bieten Rubin & Kanwisher (1985) jedoch eine alternative Erklärung der Experimente Chens an.

2) Tsotsos (1990) diskutiert Randbedingungen, welche für die Architektur des menschlichen visuellen Systems gelten müssen und leitet

Vorhersagen ab, basierend auf elementaren Schaltzeiten, der Anzahl der Neurone, der Anzahl der Verbindungen und einigen anderen bekannten Gehirneigenschaften. Falls eine der von Tsotsos genannten Architekturbeschränkungen und Vorhersagen empirisch widerlegt wird und sein Modell konsistent ist, fällt die von ihm angenommene Informationsverarbeitungshypothese für die Wahrnehmung.

Die Stärke der zweiten Grundannahme der Kognitionswissenschaften, die Annahme mentaler Repräsentationen, liegt insbesondere darin, daß sie zu einfacheren Modellen führt als andere Erklärungsversuche.

Ein typisches Beispiel hierfür sind die Ergebnisse von Stigler (1984): Es kann empirisch belegt werden, daß amerikanische und chinesische Schulkinder beim Kopfrechnen systematisch unterschiedliche Antwortzeiten haben und unterschiedliche Fehler machen. Erklärt werden die Ergebnisse damit, daß unterschiedliche mentale Repräsentationen benutzt werden. Während amerikanische Kinder klassische 'westliche' Multiplikationsalgorithmen verwenden, lernen chinesische Kinder das Rechnen mit dem Abacus. Beim Kopfrechnen finden offenbar Vorstellungen Verwendung, welche das Papier und Bleistift- bzw. das Abacusrechnen simulieren, und damit dem gleichen Zeitverhalten und den gleichen Fehlerquellen unterliegen. Andere Erklärungsversuche sind bisher unterlegen, weil sie erheblich komplizierter sind und die empirischen Ergebnisse nicht so gut erklären können.

4. Zusammenfassung

Wir halten die kognitive Theorie von Winograd/Flores für nicht überzeugend. Die wesentlichen Schwachpunkte sind die fehlende empirische Tragweite und die mangelhafte Möglichkeit der Falsifikation der empirischen Komponenten der Theorie. Die Aussagen von Winograd/Flores haben zwar eine gewisse interne Konsistenz, sie beziehen sich aber fast nicht auf die zahlreich vorhandenen empirischen Ergebnisse, z.B. aus der kognitiven Psychologie.

Winograd/Flores formulieren in ihrem Buch Kritik an der bestehenden kognitionswissenschaftlichen Theorie, ohne jedoch eine wirkliche Alternative, die den von ihnen formulierten Ansprüchen genügen würde, vorstellen zu können. Sie bemühen sich ihre Theorie der Kognition als eine auf Erfahrung beruhende aber dennoch die Erfahrung transzendierende Theorie zu präsentieren. Dieses Vorhaben muß als gescheitert angesehen werden. Die philosophischen Ansichten, die Winograd/Flores zur Unterstützung heranziehen, können die ihnen zugedachte Aufgabe nicht erfüllen. Zugleich läßt sich ein Teil derjenigen Gedanken Heideggers, die Winograd/Flores zur Stützung ihrer Theorie verwenden, in rationalistisch basierten Theorien wiederfinden. Diese Gedanken sind daher nicht in der Lage eine Theorie gegenüber einer anderen auszuzeichnen. Der andere Teil entzieht sich zumindest bei der Anwendung der von uns angegebenen Kriterien einer *sinnvollen* Interpretation. Die Erkenntnisse der Sprechakttheorie sind sicherlich wichtig, aber auch sie lassen sich in bestehende Theorien der Kognition integrieren. Eine weitere wichtige Basis von Winograd/Flores, der radikale Konstruktivismus Maturanas, ist u.E. inkonsistent und damit nicht haltbar.

Wir sehen den kognitionswissenschaftlichen Ansatz als erheblich erfolgversprechender an, denn er verfügt über eine starke empirische Verankerung, seine empirischen Komponenten können falsifiziert werden, und er erlaubt eine Vielzahl von überprüfbaren Vorhersagen. Dies bedeutet allerdings nicht, daß innerhalb der Kognitionswissenschaften gegenwärtig alle geistigen Phänomene erklärt werden können, die kognitionswissenschaftliche Theorie hat sicherlich nicht die gleiche empirische Stützung wie z.B. viele Teilgebiete der Physik. Der Ansatz dient vielmehr als Leitlinie für weitere Forschungen.

Wir hielten es deshalb gegenwärtig für einen erheblichen Rückschritt, wenn die Theorie von Winograd/Flores die kognitionswissenschaftliche Theorie verdrängen würde.

Winograd/Flores präsentieren einige wichtige Einsichten, z.B. über die Bedeutung des Kontextes und der sozialen Einbindung von Aktionen und

Handlungen und über Gefahren bei der Übertragung von Verantwortung an Computersysteme. Diese Einsichten machen aber ihr diffuses kognitives Modell nicht notwendig. Ganz im Gegenteil, sie sind auch mit dem kognitionswissenschaftlichen Forschungsprogramm verträglich und werden von dessen Anhängern überwiegend geteilt. Die große Popularität der englischsprachigen Ausgabe des Buches von Winograd/Flores erklären wir uns mit der Zustimmung zu den genannten Einsichten, welche hauptsächlich im dritten Teil des Buches zur Sprache kommen. Aber daraus auf die Plausibilität ihres kognitiven Modells zu folgern, ist ein logischer Fehlschluß.

Wir stimmen mit Winograd/Flores darin überein, daß die KI sich ihrer philosophischen Grundlagen mehr bewußt werden sollte, allerdings lehnen wir die Vorgehensweise von Winograd/Flores ab, sich passende Bruchstücke aus verschiedenen Theorien "auszuleihen", ohne sicherzustellen, daß sie ein kohärentes Gedankensystem ergeben. Darüber hinaus vermissen wir bei Winograd/Flores eine angemessene wissenschaftstheoretische Auseinandersetzung mit den Grundannahmen und den Einzelhypothesen der Theorie.

Einen letzten Punkt, den wir nicht unerwähnt lassen wollen, betrifft Maturanas Erkenntnis, daß es keinen unabhängigen, objektiven Beobachterstandpunkt gibt. Schon W. V. O. Quine wählte als Motto für sein Buch *Word and Object* (1960) ein Zitat Otto Neuraths, das u.E. sehr deutlich zeigt, daß auch schon Vertreter einer rationalistischen Denkweise dieses Problem erkannt haben:

"Wie Schiffer sind wir, die ihr Schiff auf offener See umbauen müssen, ohne es jemals in einem Dock zerlegen und aus besten Bestandteilen neu errichten zu können."

5. Anmerkungen zum Nachwort der deutschen Ausgabe und zur Übersetzung des Buches von Winograd/Flores

5.1 Das Nachwort

Wolfgang Coy, der Verfasser des Nachwortes und Vorsitzender der Fachgruppe 1.7: *KI und Gesellschaft* im Fachbereich 1 *Künstliche Intelligenz* der Gesellschaft für Informatik, präsentiert in seinem enthusiastischen Nachwort zuerst einmal die biographischen Daten von Terry Winograd und Fernando Flores. Hierbei läßt er natürlich Winograds Wandlung vom Saulus – Autor des Buches *Understanding Natural Language* – zum Paulus nicht unerwähnt. Und auch er erliegt u.E., wie Winograd/Flores dem Irrtum, daß Intelligenz eine Eigenschaft ist, die ausschließlich Menschen zukommt. Daher geht auch er explizit und implizit davon aus, daß die KI darauf abzielt, künstliche Menschen zu schaffen. Andererseits unterstellt er dem Zweig der KI, den Searle als *starke KI* bezeichnet, daß dort kognitive Fähigkeiten simuliert werden sollen. Dies trifft jedoch nicht zu, da dieser Ansatz gerade davon ausgeht, daß Informationsverarbeitung Symbolmanipulation ist und nicht nur mit Hilfe von Symbolmanipulation simuliert werden kann.

Auch zu Coys Darstellung des Buches und der darin vorgetragenen Gedanken möchten wir noch einige kurze Bemerkungen machen. Coy schließt sich der kritischen Haltung gegenüber der Korrespondenztheorie der Wahrheit an und führt als Protagonisten einer solchen Theorie u.a. Frege mit seiner *Begriffsschrift*, Whitehead & Russell mit ihren *Principia Mathematica* und Ludwig Wittgenstein mit seinem *Tractatus logico-philosophicus* an. Hier können wir uns Coys Ausführungen nicht anschließen. Ganz im Gegenteil; Frege war ein ausgesprochener Gegner der Korrespondenztheorie (vgl. Lechner ersch.). In seiner Arbeit *Der Gedanke* diskutiert Frege verschiedene Möglichkeiten, den Begriff *wahr* zu definieren. Hierbei untersucht er auch die Möglichkeit, daß das Wahrsein in der Übereinstimmung einer Vorstellung und der Wirklichkeit besteht. Dies führt jedoch zu einem infiniten Regreß, und

Frege kommt daher zu dem Schluß, daß der Versuch, "die Wahrheit als eine Übereinstimmung zu erklären", scheitert. (Frege 1918: 60)

Auch die *Principia Mathematica* von Whitehead & Russell sagen nichts über die Welt oder den Zusammenhang von Sätzen und der Welt aus, und die Autoren beanspruchen dies auch gar nicht. Die beiden Autoren betreiben die Zurückführung der Mathematik auf die Logik und ihre wenigen, grundlegenden Axiome, sie sagen jedoch nichts über das Verhältnis von Sprache und Welt. Es trifft zwar zu, daß Russell ein Vertreter der Korrespondenztheorie der Wahrheit ist, und er diese Theorie in der Einleitung der *Principia* vertritt; dies hat jedoch nichts mit den darauf folgenden, rein logischen Ausführungen zu tun. Diese sind vollständig unabhängig davon. Russell und Whitehead haben jedoch diese Art von Mißverständnis schon vorausgesehen und im Appendix C der zweiten Auflage erklärend ausgeführt:

"It does not matter, for mathematical logic, what constitutes truth or falsehood; all that matters is that they divide into two classes according to certain rules." (Whitehead/Russell 1910-13, 1925², 1978: 402)

Die *Principia Mathematica* bzw., die in ihr enthaltenen Überlegungen basieren also in keiner Weise auf der Korrespondenztheorie der Wahrheit; die Ausführungen Coys sind daher irreführend.

Etwas anders aber doch ähnlich liegt der Fall beim *Tractatus logico-philosophicus*, der Logisch-philosophischen Abhandlung Ludwig Wittgensteins. Coy zitiert einige zentrale Passagen dieses Werkes und erweckt dadurch den Eindruck, als sei Wittgenstein ein typischer Vertreter der rationalistischen Tradition, also der Tradition, die behauptet – so Winograd/Flores und Coy –, daß alles Wesentliche durch Analyse auf sprachlich formulierbare Regeln zurückgeführt werden kann. Leider läßt er hierbei entscheidende Ausführungen aus dem Vorwort unerwähnt. Dort heißt es u.a.: "Ich bin also der Meinung, die Probleme im Wesentlichen endgültig gelöst zu haben. Und wenn ich mich nicht irre, so besteht nun der Wert dieser Arbeit darin, daß sie zeigt, wie wenig damit getan ist, daß diese Probleme gelöst sind." (Wittgenstein 1921, 1977: 8)

Wittgenstein war keinesfalls ein Hurra-Vertreter einer rationalistisch-analytischen Philosophie, sondern eher ein grübelnder, (ver-) zweifelnder Denker, immer der Tatsache eingedenk, daß die Probleme des Lebens von dieser Art Philosophie nicht zur Gänze gelöst werden. Denn nur so erklären sich die scheinbar unverständlichen Ausführungen am Ende des *Tractatus*: "Meine Sätze erläutern dadurch, daß sie der, welcher mich versteht, am Ende als unsinnig erkennt, wenn er durch sie – auf ihnen – über sie hinausgestiegen ist. (Er muß sozusagen die Leiter wegwerfen, nachdem er auf ihr hinaufgestiegen ist.) Er muß diese Sätze überwinden, dann sieht er die Welt richtig." (Wittgenstein 1921, 1977: 6.54)

Auch bei der Beschreibung der momentanen Situation der KI in Deutschland trifft Coy mit seiner Wahl der Beispiele nicht immer ins Schwarze. So präsentiert er als einen weiteren, zum Scheitern verurteilten Versuch, Wissen und Verstehen in einer formallogischen Notation symbolisch zu beschreiben, das Projekt LILOG (Linguistische und logische Methoden zum maschinellen Sprachverstehen des Deutschen) des Computergiganten IBM. Es ist zwar zutreffend, daß auch bei diesem Unternehmen Verstehen auf den Einsatz geeigneter Repräsentationen zurückgeführt werden soll. Jedoch gibt es gerade innerhalb dieses Unternehmens eine Forschergruppe – LILOG-Hamburg –, die die Repräsentation von Wissen in einer ausschließlich sprachlich-propositionalen Form für unzureichend hält. Daher werden dort die Möglichkeiten untersucht, räumliches Wissen in einer bildhaft-analogen Form zu repräsentieren. Hierbei liegt das Schwergewicht jedoch nicht nur auf einer maschinengerechten Realisierung, sondern die Befunde der kognitiven Psychologie und der Psycholinguistik bilden die Basis dieses Ansatzes (vgl. Herzog et al. 1986; Rehkämper 1987, 1990; Habel 1988; Khenkhar 1989).

Alles in allem muß man feststellen, daß dieses Nachwort die Lektüre des Buches nicht angenehmer, einfacher oder gar verständlicher macht.

5.2 Die Übersetzung

Auch die Übersetzung ist leider nicht so, wie man es sich wünscht. Zum einen greift der Übersetzer bei der Übertragung von *termini technici* nicht auf schon bestehende und akzeptierte Übersetzungen zurück; zum anderen unterlaufen ihm bei der Übersetzung sinnentstellende Fehler.

Beispiele für das erste sind zum einen die Übersetzung des Terminus 'felicity conditions' mit 'Glückensbedingungen'. Hierfür hatte sich schon seit längerem im Deutschen der Ausdruck 'Gelingensbedingungen' eingebürgert, und es ist nicht ersichtlich, warum dieser nicht übernommen wurde (z.B. Wunderlich 1976). Zum anderen wurde der Maturanasche Term 'experiential' im Deutschen durch 'empirisch' wiedergegeben (z.B. 117). Auch diese Entscheidung ist uns nicht ganz verständlich.

Auch für die problematische Übersetzung möchten wir neben den oben genannten zweisprachigen Zitaten zwei weitere Beispiele angeben. Im englischen Original findet sich auf Seite 105 in Kapitel 8.5 "Can pigs have wings?" der Satz "It is not a matter of being right or wrong, accurate or inaccurate, but rather of a pre-understanding that guides our discourse and our actions." Seine deutsche Entsprechung lautet nun: "Die Frage, gestellt in der Kapitelüberschrift ['Warum wachsen Schweinen keine Flügel?', K.R. & M.M.], läßt sich jedoch nicht mit wahr oder falsch, zutreffend oder unzutreffend erledigen, sie wird vielmehr bestimmt durch unser Vor-Verständnis, das unser Argumentieren und Handeln prägt." (178) Wir sind nicht der Ansicht, daß sich ein Übersetzer sklavisch an die Vorlage halten sollte; die Übertragung des Sinns muß im Vordergrund stehen. (Folgt man den Ausführungen von Winograd/Flores ist eine wörtliche Übersetzung auch gar nicht möglich.) Hier aber muß man den Eindruck gewinnen, daß Winograd/Flores nicht in der Lage sind, Fragen und wahrheitswertfähige Aussagen auseinander zuhalten; Aussagen können wahr oder falsch, zutreffend oder unzutreffend sein, Fragen hingegen nicht.

In einem anderen Fall legt der Übersetzer eine kausale Beziehung nahe, die sich im Original nicht finden läßt. "There are many who would call themselves behaviorists (or 'neobehaviorists') who postulate an internal

state of an organism as well as the record of inputs and outputs." (48) Im Deutschen wird daraus: "Viele würden sich selbst als Behavioristen (oder Neobehavioristen) bezeichnen, weil sie den internen Zustand eines Organismus ebenso voraussetzen wie die Möglichkeit, Ein- und Ausgänge protokollieren zu können." (87, Hervorhebung durch K.R. & M.M.) Die Argumentation wirkt im Deutschen völlig unlogisch, da Behavioristen im Normalfall gerade keine internen Zustände eines Organismus postulieren. Mit diesen Bemerkungen zur deutschen Ausgabe, die sich natürlich auf besonders plakative Beispiele beziehen, möchten wir es bewenden lassen.

Alles in allem kann man potentiellen Lesern des Buches nur raten, die englische bzw. amerikanische Originalausgabe zu erwerben, die zudem auch noch deutlich preiswerter ist (12.95\$).

Danksagung: Wir danken Barbara Becker, Ansgar Beckermann, Carola Eschenbach, Claudia Meienborn, Bernd Neumann und Gerd Strube für viele kritische Kommentare und Hinweise.

Literatur:

Aebli, Hans (1980, 1981): Denken: Das Ordnen des Tuns. Bd. I: Kognitive Aspekte der Handlungstheorie. Bd. II: Denkprozesse. Stuttgart: Klett-Cotta.

Bartlett, F. (1932): Remembering. Cambridge: CUP.

Boff, Kenneth R.; Kaufman, Lloyd; Thomas, James P. (1986), Handbook of Perception and Human Performance, Volume I + II, John Wiley and Sons 1986.

Bröker, Walter (1967²): Platos Gespräche. Frankfurt/M: Vittorio Klostermann 1985³.

Carnap, Rudolf (1966): Philosophical foundation of physics, Basic books Inc., New York 1966.

Chen, Lin (1982): Topological Structure in Visual Perception. *Science* 218

Chen, Lin (1989): Topological perception: a challenge to computational approaches to vision, in "Perspectives in Connectionism", R. Pfeiffer (Ed.), Elsevier Science 1989.

Dreyfus, Hubert (1979): What computers can't do - the limits of artificial intelligence, New York, Harper & Row Publ. (Deutsche Ausgabe: Dreyfus, Hubert (1985): Die Grenzen künstlicher Intelligenz – Was Computer nicht können. Königstein/Ts: Athenäum.).

Eschenbach, C.; Habel, Ch.; Herweg, M.; Rehkämper, K. (1990): Restriktionen für plurale Diskursanaphern. In: Felix, S./Kanngießer, S./Rickheit, G. (Hrsg.): Sprache und Wissen. Opladen: Westdeutscher Verlag.

Frege, Gottlob (1879): Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens. Halle an der Saale, Louis Nebert Verlag. (Nachdruck: Hildesheim, New York: Georg Olms Verlag 1977²)

Frege, Gottlob (1918): Der Gedanke. Beiträge zur Philosophie des deutschen Idealismus. 1 (1) Auch in: Frege, Gottlob (1966): Logische Untersuchungen. Hrsg.: Günther Patzig. Göttingen: Vandenhoeck & Ruprecht 1966.

Habel, Ch. (1988): Prozedurale Aspekte der Wegplanung, in: Helmut Schnelle, Gert Rickheit (Hrsg.): Sprache in Mensch und Computer, Westdeutscher Verlag, Wiesbaden.

Hempel, Carl Gustav (1966): Philosophy of natural science, Prentice-Hall Inc., New York 1966.

Herzog, O.; Rollinger, C.-R.; Schmitt, P.; Steffens, P.; Studer, R.; Wesche, B.; Bartsch-Spörl, B.; Günthner, F.; Habel, C.; Kanngießer, S.; Rohrer, C. (1986): Lilog - Linguistische und logische Methoden für das maschinelle Verstehen des Deutschen, Projektbeschreibung. Stuttgart: IBM.

Johnson-Laird, Philip N. (1983): Mental Models. Cambridge u.a.: Cambridge University Press.

Khenkhar, Mohammed (1989): DEPIC-2D: Eine depiktionale Komponente zur Repräsentation und Verarbeitung räumlichen Wissens. In: Metzging, D. (Hrsg.): Proc. GWAI-89. Springer: Berlin.

Kuhn, Thomas S. (1970): The structure of scientific revolution, 2. erweiterte Auflage, University of Chicago Press 1970.

Lechner, Jochen (ersch.): Zum Verhältnis von Semantik und Ontologie bei Frege. In: Michael Sukale (Hrsg.) (ersch.): Sprache, Theorie und Wirklichkeit. Bern: Peter Lang Verlag.

Mohnhaupt, Michael; Neumann, Bernd (1990): Support for an Intermediate Pictorial Representation, Kommentar zu Tsotsos (1990), erscheint in "The Behavioral and Brain Science", Cambridge University Press 1990.

Platon: Sämtliche Dialoge. herausgegeben von Otto Apelt, Hamburg: Meiner 1988. (Nachdruck der Ausgabe von 1923).

Popper, Karl (1936): Logik der Forschung. Tübingen: Mohr 1989⁹.

Quine, Willard V. O. (1960): Word and Object, MIT Press, Cambridge Mass. (Deutsche Ausgabe: Quine, Willard V. O. (1980): Wort und Gegenstand. Stuttgart: Philipp Reclam jun.).

Rehkämper, K. (1987): Mentale Bilder und Wegbedeutungen. In: Morik, K. (Hrsg.): GWAI-87 11th German Workshop on Artificial Intelligence. Berlin: Springer, 297-305 (Auch erschienen als Lilog-Report 20, Stuttgart: IBM, 1987).

Rehkämper, K. (1990): Mentale Bilder - analoge Repräsentationen. In: Freksa, Chr./Habel, Ch. (Hrsg.): Repräsentationen und Verarbeitung räumlichen Wissens. Berlin: Springer (Auch erschienen als Lilog-Report 65, Stuttgart: IBM, 1988).

Rubin, J.; Kanwisher, N. (1985): Topological Perception: wholes in an experiment, *Perception & Psychophysics*, 37, 1985.

Sanford, A.J.; Garrod, S.C. (1981): Understanding written language. Chichester et al.: John Wiley & Sons.

Schank, R.C.; Abelson, R.P. (1977): Scripts, Plans, Goals, and Understanding. Hillsdale, N.Y.: Erlbaum.

Schmidt, Siegfried J. (Hrsg.) (1987): Der Diskurs des Radikalen Konstruktivismus. Frankfurt/M.: Suhrkamp.

Searle, John R. (1969): Speech Acts. Cambridge UP; dt.: Sprechakte. Frankfurt/M.: Suhrkamp 1972.

Sneed, J. D. (1971): The Logical Structure of Mathematical Physics, Dordrecht 1971.

Stegmüller, Wolfgang (1979): Hauptströmungen der Gegenwartsphilosophie, Band II, Alfred Kröner Verlag Stuttgart, 6. erweiterte Auflage.

Stiegler, James W. (1984): "Mental abacus": the effect of abacus training on chinese children's mental calculations, *Cognitive Psychology*, 16, 1984, 145-76.

Tsotsos, John K. (1990): Analysing vision at the complexity level, erscheint in "The Behavioral and Brain Science", Cambridge University Press 1990.

Whitehead, Alfred N.; Russell, Bertrand (1910-1913): Principia Mathematica 3 Bde. Cambridge 1925² [pb. to 56* 1978].

Wendel, Hans Jürgen (1989): Wie erfunden ist die Wirklichkeit? Delfin – Eine deutsche Zeitschrift für Konstruktion, Analyse und Kritik, XII (6. Jhrg.).

Wieland, Wolfgang (1982): Platon und die Formen des Wissens. Göttingen: Vandenhoeck & Ruprecht.

Wittgenstein, Ludwig (1921): Tractatus logico-philosophicus. Frankfurt/M. Suhrkamp 1977.¹²

Wunderlich, Dieter (1976): Studien zur Sprechakttheorie, Frankfurt/M. Suhrkamp Verlag.