

# Konstruierte Illusionen

## Alexander Riegler über Terry Winograds und Fernando Flores' *Understanding Computers and Cognition*

### 1 Entstehungsbedingungen und Vorgeschichte: Die Versprechungen der Artificial Intelligence-Protagonisten

Wohl kaum eine Disziplin hat in ihrer Geschichte so viele Versprechungen gemacht und gebrochen wie die *Artificial Intelligence* (AI). So hat Herbert Simon, einer ihrer Pioniere, 1965 vorausgesagt: „machines will be capable, within twenty years, of doing any work that a man can do“ (Simon zitiert nach Boden 2008: 840). Versprechungen dieser Art sind zwar nicht in Erfüllung gegangen, haben aber der Disziplin vor allem in den 1960er und 70er Jahren enorme Projektgelder beschert. Die Ansprüche waren jedoch nicht unbegründet. Alan Newell und Herbert A. Simon hatten bereits Anfang der 1960er Jahre einen „General Problem Solver“ (Newell/Simon 1963) vorgestellt, dem unzählige Expertensysteme folgten. 1966 konnte Joseph Weizenbaum (1966) mit seinem Programm ELIZA den Eindruck erwecken, Benutzer stünden einem der menschlichen Kommunikation fähigen Computer gegenüber. (Tatsächlich zeigte das Programm vielmehr, wie simpel und stichwortorientiert menschliche Konversation manchmal sein kann). Der Doktoratsstudent Terry Winograd schließlich setzte mit seinem Programm SHRDLU der Entwicklung einen Höhepunkt, bevor die Kritiker der AI immer zahlreicher wurden.

Die Artificial Intelligence widmete sich in dieser Periode (1967–1972) den so genannten „Microworlds“ – Modellen der Welt mit stark reduzierter Anzahl darin enthaltener Objekte und Relationen. Es wurde gehofft, dass man diese Mikrowelten, sobald sie beherrschbar waren, so weit vergrößern könnte, bis sie die Komplexität der Alltagswelt und damit das Niveau des menschlichen „Common Sense“ erreichen würden. Zu solchen Szenarien zählte auch die um 1970 von Terry Winograd entwickelte Klötzchenwelt SHRDLU, die an die Probleme der Wissensrepräsentation und der Verarbeitung natürlicher Sprache herantrat. Ihr Funktionspektrum umfasste das Verstehen von auf Englisch gestellten Fragen über die derzeitige Konfiguration der Mikrowelt und das Erteilen entsprechender Antworten, das Verstehen von Befehlen zur Manipulation der Klötzchen und Zerlegung der Befehle in eine Folge ausführbarer Operationen, und schließlich das Verstehen und Beschreiben der gewählten zielgerichteten deduktiven Vorgangsweise. Wie Winograd (1972) formulierte, handelt es sich also um ein Computerprogramm, das menschliche Sprache in einem, wenn auch auf Bausteine verschiedenster Formen und Farben begrenzten Bereich, versteht.

Damit schlüpfte SHRDLU für eine gewisse Zeit in die Rolle des führenden Paradigmas in der AI-Forschung. Weshalb war es von so überragender Bedeutung? Zum einen konnte es scheinbar einem lang gehegten Wunsch des Menschen nach-

kommen, nämlich dessen Sprache zu verstehen. Zum anderen verkörperte es genau das, was Alan Newell und Herbert A. Simon (1976) in der „Physical Symbol System Hypothesis“ (PSSH) zusammenfassten, nämlich das Ziel der so genannten „Starken AI“, menschliche Intelligenz in künstlichem Substrat zu reproduzieren, sie mittels symbolverarbeitender Mechanismen zu erreichen: „A physical symbol system has the necessary and sufficient means of general intelligent action“ (ebd.: 116). Diese Idee entsprach völlig dem Zeitgeist, denn sie schien übereinzustimmen mit den Erkenntnissen der kognitiven Psychologie, die ebenfalls Intelligenz als Informationsverarbeitung betrachtete (Neisser 1967). Wie Hubert Dreyfus (2007) betont, war die Idee auch philosophisch keineswegs aus der Luft gegriffen, sondern lässt sich mit einer Reihe von philosophischen Lehren in Einklang bringen: Thomas Hobbes’ Leviathan, demzufolge Denken nichts anderes als Rechnen sei; René Descartes’ Idee mentaler Repräsentationen; die *Characteristica Universalis* von Gottfried Leibniz, die Denken auf eine Reihe von Basiselementen zurückführen möchte, also auf ein universelles Symbolsystem, in dem jedem Symbol ein Objekt zugeordnet werden kann; Gottlob Freges Formalisierung der Kantschen Behauptung, Konzepte wären nichts anderes als Regeln; und Bertrand Russells Postulierung logischer Atome als die Grundbausteine der Realität. AI-Forscher versuchten, diese philosophischen Ideen in die Praxis umzusetzen. Allerdings unterscheiden sich AI-Modelle von philosophischen Entwürfen in einem kritischen Punkt: Sie verlangen bis ins kleinste Detail ausformuliert zu werden und machen es unmöglich, über „Selbstverständlichkeiten“ hinweg zu philosophieren, denn Computer müssen Schritt für Schritt instruiert werden, was zu tun ist. Dieser Sachverhalt wurde bereits im Gründungsjahr der AI gewissermaßen als epistemologische Basis der AI festgelegt: Jeder Aspekt des Lernens oder eines beliebig anderen Merkmals von Intelligenz kann derart genau beschrieben werden, um eine Maschine bauen zu können, die ihn simuliert.<sup>1</sup>

Das zweite epistemologische Credo der AI betrifft den Begriff der „Intelligenz“ selbst. Hier erwies sich der Artikel „Computing Machinery and Intelligence“ von Alan Turing aus dem Jahr 1950 als richtungweisend. Er definierte Intelligenz als Verhaltensäquivalenz: Wenn eine Maschine genauso intelligent *handelt* wie ein Mensch, dann *ist* sie auch so intelligent wie ein Mensch. Turing ersann hierzu den später nach ihm benannten Turing-Test, der die Intelligenz eines Computers als Sprachkompetenz definiert: In einer chatähnlichen textbasierten Gesprächssituation versucht ein menschlicher Fragesteller herauszufinden, welcher von seinen beiden Gesprächspartnern ein Mensch und welcher ein Computer ist. Gelingt das nicht, hat der Computer den Test bestanden und seine Intelligenz muss als die dem Menschen ebenbürtig angesehen werden.

---

<sup>1</sup> „Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it.“ (McCarthy et al. 1955: o. S.)

Beide epistemologischen Grundsätze lassen erkennen, dass es bei der AI von Anfang an um die Entwicklung eines exakten Modells menschlicher Sprachfähigkeit ging. Die Verwendung von sprachlich definierten Symbolen im Sinne der PSSH bot sich geradezu an, und AI-Forscher gingen daran, Dinge, Situationen, Ereignisse und Expertenwissen mittels geeigneter Symbole unabhängig von Kontext und Hintergrund zu repräsentieren.<sup>2</sup> Man nahm also an, dass es eine fundamentale semantische Korrespondenz zwischen Symbol und Welt gibt, die durch einen Mechanismus erfasst und verarbeitet werden kann, so dass ihm Intelligenz zugesprochen werden kann. Damit war die AI im Kern eine mechanistische Epistemologie,<sup>3</sup> die Begriffe wie Intelligenz und Erkenntnis *computational* erklärt: Vordefinierte Ausgangs-, Zwischen- und Zielzustände formen den so genannten „symbolischen Suchraum“ („symbolic search space“, Partridge 1987: 104) und können durch geeignete (mentale) Operatoren ineinander überführt werden. Aus dieser Sicht sind Problemlösen und intelligentes Handeln im Allgemeinen nichts anderes als das Auffinden und Anwenden einer Sequenz von Operationen, die den Anfangszustand über die Zwischenzustände in den Endzustand überführen: „Aufgabenlösen besteht erst in der Generierung einer geeigneten Repräsentationsstruktur (d. h. in der Transformation von Wissen) und dann in der Anwendung geeigneter Inferenzschemata“ (Freska 1988: 156). Um beispielsweise in der Blockwelt SHRDLU einen menschlichen Befehl erfolgreich ausführen zu können, muss das Programm einen Weg finden, um aus der momentanen Konfiguration durch das Ausführen einer geeigneten Handlungssequenz die gewünschte Zielkonfiguration zu machen.<sup>4</sup> Und SHRDLU konnte dies stets erfolgreich ausführen. Wo sahen also die AI-Kritiker das Problem?

Das wohl Augenfälligeste war der überzogene Anspruch, Mikrowelten auf komplexere Situationen extrapolieren zu können. Die kombinatorische Explosion von Verarbeitungsschritten, die bei zunehmend komplexeren Systemen vonnöten sind, lassen diese rasch an die Grenzen des Machbaren stoßen.<sup>5</sup> Das sich daraus ergebende „Frame Problem“ (Dennett 1984) macht es für ein künstlich intelligentes System unmöglich, aus der unüberschaubaren Anzahl von Alternativen die jeweils

<sup>2</sup> Der so genannte „Repräsentationalismus“ nimmt an, dass materielle Objekte in mentalen Bildern repräsentiert werden. Die Korrektheit menschlichen Wissens hängt dann davon ab zu zeigen, dass die Bilder mit den externen Objekten akkurat übereinstimmen. Im Konstruktivismus wird davon ausgegangen, dass dieser Vergleich nicht möglich ist, am vehementesten etwa von Ernst von Glasersfeld (1997: 26): „Um [den Vergleich] durchzuführen, müsste man unmittelbaren Zugang zu einer Realität haben, die jenseits der eigenen Erfahrung liegt und von den eigenen ‘Bildern’ und ihren sprachlichen Darstellungen unberührt bleibt“.

<sup>3</sup> An anderer Stelle bezeichnet Terry Winograd diese Mischung von „common sense introspection, *ad hoc* programming and so-called ‘knowledge acquisition’ techniques for interviewing experts“ als „patchwork rationalism“ (Winograd 1990: 171 f.).

<sup>4</sup> Die dabei verwendete „means-ends analysis“ wurde so zum Standardrepertoire der AI.

<sup>5</sup> Ein entsprechender Evaluationsreport aus dem Jahre 1973, der dies als die Hauptursache für das Scheitern der AI angab, brachte das AI-Programm in Großbritannien beinahe zum Erliegen.

relevante herauszufinden. Für natürlich intelligente Systeme scheint das hingegen kein Problem zu sein: Menschen verhalten sich auch in komplexen Umwelten intuitiv richtig. Für AI-Kritiker wie Hubert Dreyfus war das der entscheidende Punkt zu proklamieren, dass intelligente Maschinen dies prinzipiell nie können werden. Man müsse vielmehr deutlich zwischen zwei Bereichen unterscheiden: jene, in denen Menschen intuitive Experten sind (wie etwa Schachspielen, das Erstellen medizinischer Diagnosen und vor allem sprachliche Kompetenz) und jene Bereiche, wo es keine intuitiven Experten gibt. Da in letzteren menschliche Intuition nicht greift, triumphieren hier Computersysteme, weil sie die zur Bewältigung erforderlichen Handlungssequenzen wesentlich rascher berechnen können als der menschliche Experte (vgl. Partridge 1987).

Dass vor allem Sprache zu den intuitiven Bereichen gehört, fand Winograd heraus, als sein Versuch scheiterte, in weiterer Folge eine Wissensrepräsentationssprache zu entwickeln, in der semantische Bedeutung aus syntaktischen Elementen aufgebaut werden kann.<sup>6</sup> Die kombinatorische Explosion von Entscheidungsmöglichkeiten im Suchraum machte es dem Programmierer unmöglich, alle Eventualitäten vorauszusehen, so dass leicht Fälle eintreten können, in denen das System nicht weiß, welchen Weg es einschlagen soll – ein Phänomen, das später von Terry Winograd im Anschluss an Martin Heidegger „blindness of representation“ (Winograd 1990: 180) genannt werden sollte: Die Interpretation von dekontextualisierten Symbolen obliegt dem Programmierer; vom programmierten AI-System kann es hingegen nicht geleistet werden, auch wenn Kontexte vom Programmierer weiter konditional aufgespalten werden können, so dass Alternativen feiner unterschieden werden („im Fall A tue M, im Fall B tue N usw.“). Auf diese Weise hoffte die AI-Forschung, Expertensysteme bauen zu können, die beispielsweise das medizinische Diagnosewissen von Ärzten umfassen, welche aufgrund von Symptomen (= konditionale Kontexte) eine Behandlungsdiagnose erstellen.<sup>7</sup> Es zeigte sich, dass sich in jeder Kontextsituation wieder weitere Ambiguitäten für das rein auf syntaktischer Ebene symbolverarbeitende AI-System eröffnen, so dass ein infinites Regress entsteht und die kontextuelle Bedeutung für das System auf ewig jenseitig bleiben muss – was später als „Symbol Grounding Problem“ (Harnad 1990) bekannt wurde, d. h. das Problem, Bedeutung an bedeutungslose Symbole zu knüpfen.<sup>8</sup>

Vor diesem Hintergrund von offensichtlichen Problemen, die die AI-Forscher davon abzuhalten scheinen, jemals die von ihnen versprochenen intelligenten Sys-

---

<sup>6</sup> Dies spiegelt sich u. a. in der Tatsache wider, dass sein Publikationsprojekt *Language as a Cognitive Process* nie über den ersten Teil *Syntax* (Winograd 1983) hinauskam.

<sup>7</sup> Typische Expertensysteme aus jener Zeit umfassten mehrere tausend Regeln, die die Relationen zwischen Symptomen propositional erfassten, und konnten doch nur in 75 Prozent der Fälle eine richtige Diagnose erstellen (vgl. Winograd 1990: 179).

<sup>8</sup> John R. Searle (1980) hat den fehlenden zwingenden Zusammenhang zwischen Syntax und Semantik in seinem bekannten „Chinesischen Zimmer“ illustriert.

teme zu bauen, geriet Terry Winograd, der vormalige „Shootingstar“ unter den Vertretern der symbolischen AI, der seine Karriere in der AI am heiligen Gral derselben, dem MIT, begonnen hatte, unter den argumentativen Einfluss des vehementen AI-Kritikers Hubert Dreyfus. In weiterer Folge wandte er sich von der technisch-rationalistischen Perspektive ab und der phänomenologisch-hermeneutischen zu. 1973 zog er nach Stanford und beschäftigt sich heute vornehmlich mit dem Problem der Maschine-Mensch-Schnittstelle (*human-computer interaction*, HCI), das auch im besprochenen Buch ein wichtiges Thema darstellt. Er verbrachte u. a. einige Zeit als Gastforscher bei Google, dessen Mitgründer Larry Page als Doktoratstudent an Winograds Projekt „People, Computers, and Design“ mitwirkte.

Nicht weniger ungewöhnlich verlief das Leben von Fernando Flores. Dieser wurde 1970 als 27-jähriger Ingenieur Finanzminister in der Regierung des chilenischen Präsidenten Salvador Allende. Unter Mitarbeit des Kybernetikers Stafford Beer sollte er das Computersystem Cybersyn kreieren, um die Wirtschaft Chiles in Echtzeit zu planen und zu steuern. Noch bevor diese Pläne vollständig umgesetzt werden konnten, kam es 1973 zur Machtergreifung Augusto Pinochets, die Flores drei Jahre politische Gefangenschaft einhandelte. In dieser Zeit machte er sich mit den philosophischen Werken von Humberto R. Maturana und Francisco J. Varela, Martin Heidegger, Jürgen Habermas und John R. Searle vertraut (vgl. Fisher 2009). Durch die Intervention von Amnesty International konnte er schließlich 1976 der psychologischen Folter seiner Gefangenschaft durch die Ausreise in die USA entkommen. Er begann mit dem Studium einer Reihe von Fächern, u. a. Sprachphilosophie und Management, und machte seinen PhD an der UC Berkeley unter den Brüdern Hubert und Stuart Dreyfus, sowie John R. Searle und Ann Markusen (vgl. Rubin 1999). Martin Heideggers Werk, vor allem sein Grundsatz „Sprache ist das Haus des Seins“, war für Flores von grundlegender Bedeutung. Es wurde ihm deutlich, dass die Funktion der Sprache nicht allein in der Übermittlung von Information, sondern auch von Verpflichtungen besteht. Diese Einsicht legte auch den Grundstein für die Zusammenarbeit mit Terry Winograd, mit dem er sich einig war, dass die Funktion intelligenter Computer darin bestehen müsse, nicht so sehr rational Informationen zu verarbeiten, sondern in Sprache hervorgebrachte Verpflichtungen speichern und nachverfolgen zu können. Bereits in seiner Doktorarbeit erarbeitete er eine von der Philosophie Heideggers und der seines Landsmanns Humberto R. Maturana beeinflusste Coaching- und Workflowtechnologie. Diese entwickelte er in Assoziation mit Terry Winograd weiter in der Softwarefirma Action Technologies, die sich mit Workflowanalyse, Groupware, Softwaredesign und Geschäftsprozessanalyse beschäftigte – Themen, die auch im vorliegenden Buch wiederzufinden sind. Später wurde Flores ein gefragter Managementberater, dessen Methode sich vor allem auf Sprechakte und ihre Beherrschung konzentrierte, da er davon ausging, dass die meisten menschlichen Handlungen in Konversationen ausgeführt werden. Für ihn waren Sprechakte Sprachrituale, die Vertrauen zwischen Kollegen und Kunden schaffen, Wortpraktiken, die die Sicht

auf neue Möglichkeiten erschließen. Menschen, so die Überzeugung Flores', wissen nicht, dass sie diese Möglichkeiten nicht kennen. Um diesen Bereich des Nichtwissens über das eigene Nichtwissen erschließen zu können, ist aber Vertrauen nötig. Flores' Nähe zur Philosophie von Humberto R. Maturana lässt sich besonders in einer seiner Aussprachen erkennen: „I never blamed Pinochet, or my torturers, or external circumstances. I feel 'co-responsible' for the events that took place“ (zitiert nach Rubin 1998). Dies deutet auf das konstruktivistische Verständnis von Fernando Flores hin, wonach kognitives Handeln im Hervorbringen einer Welt besteht, und nicht im passiven Repräsentieren von Objekten und Ereignissen (siehe auch Francisco J. Varela in Pörksen 2002: 115).

Die Lebensgeschichte der Autoren macht deutlich, dass ihr gemeinsames Buch keineswegs ein zufälliges Produkt ist, sondern aus dem Zusammentreffen von philosophischer, ökonomischer und politischer Einsicht hervorgegangen ist. Ausgehend von den Versprechungen der AI führte sie zu der ernüchternden, aber auch neue Chancen bietenden Erkenntnis, dass Sprache und Kognition keine mechanistisch-realistisch zu erklärenden Phänomene sind, sondern einer konstruktivistischen Herangehensweise bedürfen.

## **2 *Understanding Computers and Cognition* als Schlüsselwerk des Konstruktivismus**

Auf den einfachsten Nenner gebracht ist das Buch *Understanding Computers and Cognition* (Winograd/Flores 1986) ein Sturmloch gegen die „rationalistische Tradition“ westlicher Wissenschaftsauffassungen und -praktiken, deren Fehlleistung nirgends deutlicher zu Tage tritt als in der Artificial Intelligence und der Kognitionswissenschaft. Das Buch beinhaltet mehrere Kerngedanken. Zum einen baut es auf der Philosophie Humberto R. Maturanas und Francisco J. Varelans auf, zum anderen benützt es die Philosophien Martin Heideggers und Hans-Georg Gadammers, um die Unmöglichkeit der Explikation dessen, was wir als Intelligenz und Verstehen bezeichnen, aufzuzeigen. Weitere Anleihen werden bei der Sprechakttheorie John L. Austins und John R. Searles genommen, die allerdings im Folgenden nicht weiter aufgegriffen werden sollen.

Der Grund für den von Terry Winograd und Fernando Flores an den Tag gelegten Sturmloch liegt darin, dass die zeitgenössische Ansicht über Computer und deren Einfluss auf die Gesellschaft durch die rationalistische Tradition geprägt ist, die einer Neubewertung bedarf und als Verstehensgrundlage infrage gestellt werden sollte.<sup>9</sup> Folglich wollen Winograd und Flores in ihrem Buch die Undurchsichtigkeit

---

<sup>9</sup> Im Original: „current thinking about computers and their impact on society has been shaped by a rationalistic tradition that needs to be re-examined and challenged as a source of understanding“ (Winograd/Flores 1986: 14)

der rationalistischen Tradition und die Blindheit, die sie hervorbringt,<sup>10</sup> aufzeigen. Was charakterisiert nun die „rationalistische Tradition“? In erster Linie verbinden die Autoren diese mit dekontextualisierten Entscheidungsfindungen. Sie argumentieren, dass die (wie in der Einleitung beschriebene) Vorstellung von einem allgemein definierbaren symbolischen Suchraum mit allgemeingültigen Verzweigungen nur in sehr idealisierten Situationen sinnvoll ist. Als Modell für menschliche Kognition ist sie unbrauchbar, denn Kognition arbeitet mit individuell höchst unterschiedlichem Hintergrundwissen, das größtenteils unexplizierbar ist. Die Autoren streiten nicht ab, dass es formale Situationen geben mag, in denen rationales Vorgehen möglich ist, und für die eine symbolische Repräsentationsstruktur mit darauf arbeitenden, genau definierbaren Regeln gefunden werden kann, so dass durch Anwendung dieser Regeln Handlungen festgelegt werden können, um ein gewünschtes Ziel zu erreichen. In den meisten Bereichen menschlicher Kognition ist dies aber nicht möglich, da die Relevanz von Alternativen unbestimmbar bleibt: Relevanz ergibt sich immer aus einer Vororientierung vor einem Hintergrund (Winograd/Flores 1986: 149).<sup>11</sup> In anderen Worten, Vorurteile sind der Kognition höchst dienlich, ganz im Sinne Gadamers, für den Vorurteile „Bedingungen dafür sind, dass wir etwas erfahren, dass uns das, was uns begegnet, etwas sagt“ (zitiert nach Winograd/Flores 1989: 257). Diese Ideen sind für den radikalen Konstruktivismen keineswegs neu und lassen sich beispielsweise bei Ernst von Glasersfeld finden, der im Anschluss an Jean Piaget an vielen Stellen die Wichtigkeit des Zusammenspiels von Assimilation (Eingliederung neuer Erfahrungen) und Akkommodation (Anpassung der kognitiven Struktur, um die Eingliederung einer neuen Erfahrung zu ermöglichen) betont.

Um ihr Argument zu untermauern, greifen Winograd und Flores auf das aus der Theorie Humberto R. Maturanas bekannte Konzept der Strukturkoppelung zwischen autopoietischen (d. h. sich selbsterzeugenden) Systemen und ihrem Milieu zurück: Durch einen Prozess der reziproken strukturellen Störung lösen Organismus und Milieu beim jeweils anderen Strukturveränderungen aus. Die wechselseitigen Strukturveränderungen zweier Organismen bzw. eines Organismus und seines Milieus werden als „strukturelle Koppelung“ bezeichnet.

Ferner übernehmen sie von Maturana die Behauptung, dass sich „der Bereich der Interaktion, in dem Verhalten beobachtet wird, und der strukturelle Bereich, in dem *Strukturveränderungen* stattfinden“ (Maturana 1985: 20), einander nicht überlappen. Die Orthogonalität dieser Bereiche findet sich in zwei zentralen Konzepten Heideggers, für den die in unseren alltäglichen Aktivitäten involvierten Gegenstände („Zeug“) unserer bewussten Aufmerksamkeit entbehren. Heidegger spricht hier

---

<sup>10</sup> „[T]he non-obviousness of the rationalist tradition and [...] the blindness that it generates“ (Winograd/Flores 1986: 17).

<sup>11</sup> „Relevance always comes from a pre-orientation within a background“ (Winograd/Flores 1986: 149).

von der „Zuhandenheit“ der Dinge, wenn wir uns in einem Handlungsfluss befinden, in dem die Dinge gleichsam repräsentational unsichtbar werden.<sup>12</sup> Sobald aber eine unerwartete Störung, ein Versagen auftritt, kann dies zum Kollaps<sup>13</sup> führen, der aus der Zuhandenheit eine „Vorhandenheit“ macht, wodurch die Transparenz verfliegt und reflektiertes Handeln das eben noch unreflektierte ablöst.<sup>14</sup> Aus dem Unterschied zwischen Zu- und Vorhandenheit erklärt sich die prinzipielle Unmöglichkeit, das Hintergrundwissen, das wir verwenden, um unreflektierten intuitiven Tätigkeiten nachzugehen, vollständig zu explizieren. Michael Polanyi (1966) hat das unsichtbare Hintergrundwissen als „tacit knowledge“ bezeichnet. Es kann nicht in einer Sammlung von Propositionen erfasst werden. Jedem rationalistischen AI-Systementwurf liegt aber die Notwendigkeit zugrunde, Sachverhalte propositional angeben zu können, d. h. die Repräsentation in Symbolen muss möglich sein. Sie müssen gewissermaßen rekonstruiert werden, woraus sich eine „Blindheit“ ergibt, wie sie für symbolische AI-Systeme kennzeichnend ist: Hat der Programmierer eine Alternative im Suchraum nicht bedacht, so wird das System an dieser Alternative scheitern. In diesen Situationen ist ein AI-System hilflos, denn sein Programmierer hat bei aller Sorgfalt nicht alle Eventualitäten voraussehen können, und das nicht nur, weil es derer viele gibt, sondern eben weil prinzipiell nicht alle möglichen Verzweigungen im Such- und Entscheidungsraum *a priori* explizierbar sind. Ein AI-System bauen zu können, dessen kognitive Fähigkeiten und dessen komplexes Hintergrund- bzw. Alltagswissen (Common Sense) denen des Menschen gleichkommen, muss daher eine Illusion bleiben. Die Problemräume, die bei der Programmierung von Schachprogrammen in der AI definiert werden, und die, die dem Common-Sense-Wissen zugrunde liegen, sind notwendigerweise qualitativ verschieden. Francisco J. Varela (1990) hat diesen Sachverhalt anhand einer Grafik schön illustriert, wo der genau „kristallartig“ definierte und relativ kleine Raum eines Schachprogramms, in dem sämtliche Relationen exakt spezifizierbar sind, dem chaotisch anmutenden „Haufen“ des Alltagswissens mit seinen ungenauen Definitionen gegenübersteht (siehe Abbildung 1).

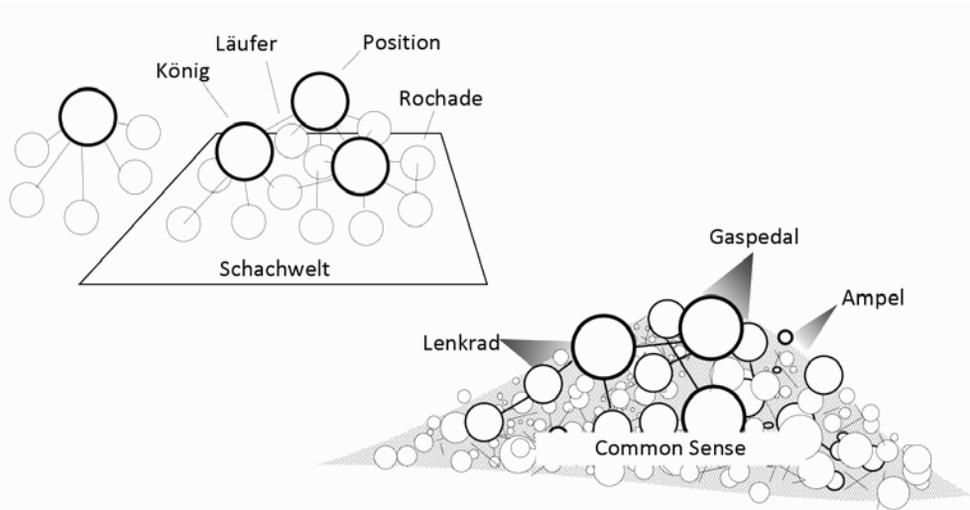
---

<sup>12</sup> Auch Mihály Csíkszentmihályi (1999) spricht von unreflektiertem „Flow“-Erlebnis, einem Schaffensrausch, von dem man ergriffen wird, wenn man in einer Aktivität völlig aufgeht.

<sup>13</sup> Die Autoren verwenden hier „breakdown“, das in seiner direkten deutschen Entsprechung bei Heidegger nicht vorkommt (vgl. Winograd/Flores 1989: 315, Anmerkung 5).

<sup>14</sup> Dieses Thema findet sich auch bei der derzeitigen Diskussion über den „Freien Willen“, in der das Spektrum zwischen Automatismen und bewusstem Willen Gegenstand der Untersuchung ist, vgl. etwa Wegner (2002).

Abb. 1: Repräsentationsräume (nach Varela 1990: 95)



In dieser qualitativen Verschiedenheit begründet sich auch der Verlust der Illusion, Mikrowelten wie jene in Winograds Programm SHRDLU bis hin zur menschlichen Alltagswelt erweitern zu können.

Die Sachlage dürfte sich allerdings weniger dramatisch gestalten als Terry Winograd und Fernando Flores annahmen. Ein Ausweg deutet sich beispielweise bei Ulric Neisser (1976) und seinem Konzept der schemageleiteten Informationsaufnahme (*schemata-controlled information pickup*) an, dem zufolge die Aufnahme von Sinneseindrücken die Wahrnehmung antizipierenden Stellen in kognitiven Schemata verändert, wodurch die Aufnahmebedingungen für weitere Eindrücke modifiziert werden und sich damit ein selbstreferenzieller Zyklus von Aufnahme und Strukturänderung einstellt. Dass der Neisser'sche Wahrnehmungszyklus nicht in Konflikt gerät mit der Kritik von Winograd und Flores, liegt daran, dass diese ihre Kritik auf performative Aspekte symbolverarbeitender AI-Systeme beschränken und Lernaspekte in ungenügendem Maße berücksichtigen. Damit tragen sie der Möglichkeit nicht Rechnung, dass Systeme durch eine Lernkomponente gewissermaßen über das, was ihnen durch den Programmierer an repräsentationellen Strukturen von außen aufgezwungen wurde, hinausgehen.<sup>15</sup> Es vergehen auch an die 100 Seiten, bevor sie sich der Frage „What about learning and evolution?“ zuwenden (Winograd/Flores 1986: 100). Ihre Antwort teilt Lernen in drei Kategorien ein: 1. Das „Justieren von Parametern“, dem sie jegliche innovative Kapazität abspre-

<sup>15</sup> Hier wird einmal mehr deutlich, dass Winograd und Flores in Anlehnung an Maturana die Fremdinstruktion intelligenter Systeme ausschließen.

chen; 2. die psychologisch motivierte kombinatorische Formung von Konzepten, die aber ein System nicht aus den Grenzen vorgegebener Elemente befreien kann; und 3. strukturelle Evolution, der sie bescheinigen, mit der Idee der strukturellen Koppelung kompatibel zu sein (und wohl auch mit Neissers schemageleiteten Informationsaufnahme; vgl. Riegler 2007). Allerdings fügen sie an, dass evolutionäre Vorgänge in technischen Systemen eine Illusion bleiben müssen, hat doch die natürliche Evolution Jahrmilliarden gebraucht, um das intelligente System Mensch hervorzubringen. Gerade dieser Kritikpunkt scheint aber mittlerweile hinfällig geworden zu sein, da kurz nach dem Erscheinen des Buches bereits Schritte in eine Richtung unternommen wurden, die heute als Genetische Algorithmen (Goldberg 1989) bzw. Genetisches Programmieren (Koza 1992) bekannt ist. Zwar beziehen sich Winograd und Flores auf das 1966 erschienene Buch *Artificial Intelligence Through Simulated Evolution* (Fogel/Owens/Walsh 1996), sind aber ignorant gegenüber den erfolgreichen Versuchen Ingo Rechenbergs (1973), mittels evolutionärer Methoden technische Systeme zu optimieren. Bereits zuvor – in den 1950ern – arbeitete Gordon Pask an elektromechanischen Maschinen mit evolvierenden sensorischen Fähigkeiten, die dadurch ganz im Sinne der Kritik von Winograd und Flores adaptiv ihre eigenen Sensoren konstruieren (Cariani 1993).

### 3 Rezeption und Wirkung

Das Buch wurde in der AI-Gemeinde genauso wenig mit offenen Armen willkommen geheißen wie schon Hubert Dreyfus' Kritik rund zwei Jahrzehnte davor (Dreyfus 1965, 1972). Dreyfus hatte als Philosoph argumentiert, dem vorgeworfen wurde, keine Ahnung von technischen Einzelheiten zu haben („his insight into programming is so poor that he classifies as impossible programs a beginner could write“ schrieb der AI-Forscher Seymour Papert über Dreyfus, zitiert nach Boden 2006: 845). Was bei Terry Winograd und Fernando Flores nun schlimmer wog, war die Tatsache, dass einer der Autoren des ketzerischen Werkes ein angesehener Vertreter der symbolischen AI war. Ähnlich wie der späte Ludwig Wittgenstein von seinen ehemaligen Weggefährten in Cambridge nicht mehr verstanden wurde, so war auch der vom Saulus zum Paulus gewandelte Winograd für viele eine große Enttäuschung. Die Reaktion in der AI-Fachwelt war eindeutig: Weitgehende Ignoranz. Abgesehen von einer vier Artikel umfassenden Kritik in der führenden Zeitschrift *Artificial Intelligence* (Clancey/Smoliar/Stefik 1994)<sup>16</sup> in dem auf die Veröffentlichung folgenden Jahr, wurde dem Buch vom Kern der symbolischen „good old-fashioned AI“ keine große Beachtung mehr geschenkt. So wird etwa im *Hand-*

---

<sup>16</sup> Die vier Buchbesprechungen erschienen zuerst 1987 in *Artificial Intelligence*, Jahrgang 31, S. 213–261 und danach gesammelt in Clancey/Smoliar/Stefik (1994: 173–221).

*book of Knowledge Representation* (Harmelen/Lifshitz/Porter 2008) der Name Terry Winograd kein einziges Mal erwähnt, geschweige denn das Buch.

Wie Michael Mohnhaupt und Klaus Rehkämper (1990) in ihrer Kritik zur deutschen Übersetzung des Buches (Winograd/Flores 1989) anmerken, sei ein großer Schwachpunkt des Buches in seiner geringen empirischen Tragweite zu sehen, wodurch sich die getroffenen Behauptungen auch der Möglichkeit der Falsifikationen weitgehend entzögen. Auch ist den Rezensenten das Buch zu fragmentarisch, weil es eine Reihe neuer, exotisch anmutender, aber letzten Endes loser Begriffe aus verschiedenen Philosophien zusammenträgt (wobei sie auch nicht mit Polemik gegenüber Heidegger sparen, an dessen exzentrischem Sprachgebrauch sie sich eingehend belustigen), so dass keine „Brückenprinzipien“ erkennbar sind. Schließlich erscheint den Kritikern das gewählte Format zu destruktiv: Zwar werden allerlei Missstände angeklagt, aber keine handfesten Alternativen geboten. Ihr Gesamturteil kann dann auch symptomatisch für die Rezeption des Werkes in der AI-Gemeinde schlechthin gelten: „Wir hielten es deshalb gegenwärtig für einen erheblichen Rückschritt, wenn die Theorie von Winograd und Flores die kognitionswissenschaftliche Theorie verdrängen würde“ (Mohnhaupt/Rehkämper 1990: 43). Zu radikal (oder auch bloß zu sehr die eigenen Pfründe bedrohend) scheint die Revolution gegen Repräsentationalismus und rationalistische Informationsverarbeitung in den Augen vieler angemutet zu haben, als dass sie den Ausführungen ein genügend hohes Maß an „commitment“ (das Winograd und Flores als zentralen Aspekt menschlicher Kommunikation ansehen) hätten angedeihen lassen können.

Im Bereich der Mensch-Maschine-Schnittstellen-Forschung, für die die Autoren des Buches neue Impulse zu geben erhofften, war die Resonanz größer. Allerdings wird das Buch in Retrospektiven eher für seine Rolle in der „language-action“- versus „situated action“-Kontroverse hervorgehoben denn für etwaige konstruktivistische Impulse. Die Reinterpretation des von Winograd und Flores vorgeschlagenen Kommunikationsmodells aus der Sicht des Konstruktivismus wurde daher versäumt. So betont beispielsweise Giorgio De Michelis (2008) in seinem Rückblick, dass das Buch viele Forscher und Designer im Bereich des *computer-supported cooperative work* zum Studium der Kontinentalphilosophen Edmund Husserl, Martin Heidegger, Hans-Georg Gadamer und Jean-Luc Nancy veranlasst hat – von konstruktivistischer Lektüre ist keine Rede. In vielen der eher praxisorientierten Bücher zum Thema Webnutzung und -technologien werden Winograd und Flores aber nicht genannt – zu abstrakt scheinen die revolutionären Ideen für die Umsetzung für das Internet und für Web 2.0 zu sein, die zum Zeitpunkt der Publikation noch auf sich warten ließen. Gerade aber das Web und seine kommunikativen und kollaborativen Aspekte sind jener paradigmatische Fall, den Winograd und Flores als die realisierbare soziale Dimension von Computertechnologie propagierten. Hier gehen die Autoren von Maturanas Begriff des konsensuellen Bereiches aus, der durch die wechselseitige Orientierung in den beteiligten Systemen charakterisiert ist: Durch die strukturelle Koppelung zweier Systeme entsteht

ein Bereich des koordinierten Verhaltens, der durch die Struktur der Systeme und ihre Entstehungsgeschichte festgelegt ist. Dadurch entstehen *commitments* – Versprechungen, Verpflichtungen und Bindungen, die mit sprachlichen Handlungen einhergehen. Die Sichtweise der rationalistischen Tradition, die Sprache als einen Informationsträger auffasst, der sich auf dekontextualisierte Gegebenheiten bezieht, ist notwendigerweise eine armselige Perspektive, die die verbindlichen Aspekte von Sprache außer Acht lässt. Da Computer keine Verbindlichkeiten eingehen können, sind sie auch nicht in der Lage, an menschlicher Sprache teilzunehmen. Ihre Rolle muss darauf beschränkt bleiben, *intelligent tools* zu sein – wie der in Winograd und Flores vorgestellte *coordinator* –, die Strukturen zur Verfügung stellen, welche den kommunikativen Austausch zwischen Menschen möglich machen und erleichtern. Mit anderen Worten: die Fähigkeiten dieser Tools bleiben auf die Manipulation syntaktischer Aspekte begrenzt und können sich nicht auf semantische Inhalte beziehen.<sup>17</sup>

Im Bereich der rezenten AI-Forschung sind viele Konzepte des Buches (ob nun direkt oder auf Umwegen sei dahingestellt) wieder aufgetaucht. Die AI ging in den Jahren, die der Publikation folgten, in den bekannten „AI Winter“. In dieser Zeit wurde das Budget für die AI-Forschung gesenkt, die zu hohen Erwartungen und die sich daraus ergebenden Versprechungen zurückgeschraubt, und der ehemals umfassende Anspruch der AI, eine globale, dem Menschen ähnliche künstliche Intelligenz schaffen zu können, von vielen aufgegeben. Die Voraussagen der Autoren, wie auch bereits die von Hubert Dreyfus und Joseph Weizenbaum in den Jahrzehnten davor, bewahrheiteten sich. Die Lösung des Common-Sense-Problems wurde nie erreicht. Stattdessen kam es zur Umlenkung der Forschungsschwerpunkte auf die Aspekte von Intelligenz, die von den Pionieren als trivial angesehen wurden, wie z. B. Navigation in unbekanntem Gelände und perzeptive Erkennungsleistungen, die mit dem symbolischen Ansatz nicht erbracht werden konnten. Heidegger-Kenner Hubert Dreyfus hatte das bereits in seiner 1965 erschienenen Polemik „Alchemy and Artificial Intelligence“ erkannt: „the body plays a crucial role in making possible intelligent behavior“ (zitiert nach Boden 2008: 839). Zu dieser neuen „embodied“ AI gehören etwa die Arbeiten Rodney Brooks, der zu Beginn der 1990er Jahre ausgerechnet an der ehemaligen Hochburg des symbolischen Ansatzes, dem MIT, Roboter zu bauen begann, die gänzlich ohne symbolische Repräsentation und Planung kognitive Fertigkeiten zuwege brachten (Brooks 1991).<sup>18</sup> Die beiden Begriffe „embeddedness“ und „embodiment“, die seitdem verstärkt in der Literatur ihre Runden machen (Riegler 2002), und die für die Erbauung intelligenter Systeme nun unerlässlich zu sein scheinen, stehen in engem Zusammenhang

---

<sup>17</sup> Hier wird die Parallele zum klötzchenmanipulierenden SHRDLU aus Winograds Frühzeit deutlich.

<sup>18</sup> Allerdings distanzierte sich Rodney Brooks von der Idee, von der Philosophie Heideggers beeinflusst worden zu sein: „It Isn't German Philosophy“ (Brooks 1991: 155).

mit den Ausführungen von Winograd und Flores und deren Interpretation der zentralen Begriffe Heideggers wie „Dasein“ und „Zuhandensein“. Auch weitere und jüngere Entwicklungen in der AI und Kognitionswissenschaft – wie etwa die *Enactive Cognitive Science*, die Francisco Varela mitbegründete (Varela/Thomson/Rosch 1991), oder die dynamische Kognitionswissenschaft (Van Gelder 1998), die vom AI-Forscher Randall Beer, einem begeisterten und „bekehrten“ Leser von Winograd und Flores, mitgetragen wird – heben sich von der in Winograd und Flores geäußerten Kritik ab.

Ein letzter Kritikpunkt am Buch ist die Feststellung, dass es mit strengen Begründungen und ausgiebigen Beispielen geizt. Die Intention der Autoren ist es ja gerade nicht, eine rationalistische Analyse zu betreiben, sondern einen bestehenden Missstand aufzuzeigen und aus den Angeln zu heben, und die Unbrauchbarkeit desselben dem Leser vor Augen zu führen. Dabei wird leider, wie viele Kritiker des Buches anmerken, die Allgemeingültigkeit der Behauptungen stark untergraben, weil sie den Eindruck hinterlassen, nur auf bestimmte Ausnahmefälle anwendbar zu sein oder nur quantitative denn qualitative Probleme zu berühren. Gerade für den technisch orientierten AI-Forscher, den die Autoren gerne mitreißen möchten, stellt sich oft die Frage, ob gewisse behauptete Unmöglichkeiten nicht durch ein Mehr an technologischem (anstelle von revolutionär-philosophischem) Aufwand zu überwinden wären. Das Buch ist damit eher ein Patchwork nicht immer schlüssig ineinander überführender Argumente, eine Sammlung von zum Teil äußerst wichtigen Erkenntnissen aus den Bereichen Konstruktivismus, Phänomenologie und Hermeneutik.

Auf jeden Fall haben die Autoren recht behalten mit ihrer Umdefinierung von Artificial Intelligence. Den Schlag gegen symbolische AI-Systeme, die aufgrund der fehlenden intuitiven Prä-Orientierung zu keinerlei intelligenter Leistung im menschlichen Sinne fähig sind, nutzen die Autoren geschickt, um darauf hinzuweisen, was ihrer Einschätzung zufolge mit Computersystemen sehr wohl im Bereich des Machbaren liegt. Sie schlagen vor, dass wir die AI-Technologie gebrauchen können, um Entscheidungshilfen zu konstruieren, die uns helfen, vordefinierte Entscheidungsräume auszuloten und Strukturen zu konstruieren, mit deren Hilfe nicht-intuitive Bereiche besser bewältigt werden können. Wie Winograd in einem der Veröffentlichung des Buches vorangehenden Kommentars meinte: „The result may not be ‘intelligent machine’ but intelligent uses of machine capabilities“ (Winograd zitiert nach Bobrow/Hayes 1985: 395). Die kollektive Intelligenz, die durch Web-einrichtungen wie Blogs, Wikis, Facebook usw. ermöglicht wird, wurde weder von den frühen AI-Forschern noch von Science-Fiction-Autoren, wohl aber in gewisser Weise von Winograd und Flores vorausgesagt.

Im Gegensatz zu der stürmischen Entwicklung im Bereich der computergestützten Sozialwerkzeuge ergeben sich für den Konstruktivismus aus dem Buch keine direkten Weiterentwicklungen. Es muss aber als Versuch gesehen werden, die Konzepte und Theorien Maturanas und Varelas auch auf technische Systeme

anzuwenden. Weiters sollte auch der Wert nicht unterschätzt werden, der im Vergleich des Konstruktivismus mit der Phänomenologie Heideggers und Gadamers (und interessanterweise der Sprechakttheorie Searles, welcher dem Konstruktivismus nicht notwendigerweise freundlich gegenüber steht, siehe Searle 1997) liegt. Dieser Vergleich und die Zusammenlegung sind aber im Buch unvollständig geblieben. Vielmehr haben die Autoren aus den jeweiligen Theorien Maturanas, Heideggers und Searles diejenigen Aspekte herausgepflückt, die ihnen zur Untermauerung ihrer „anti-illusionistischen“ (Winograd/Flores 1994) Kritik dienlich waren.

#### 4 Schlussbemerkungen

Hat das Buch letzten Endes dazu geführt, den leeren Versprechungen und Illusionen der Artificial Intelligence ein Ende zu bereiten? Wohl kaum. Seit dem Erscheinen des Buches sind viele neue Spieler auf dem Jahrmarkt der AI erschienen, die *embodied* oder evolutionäre Ansätze verfolgen, während die alten weiterhin fröhlich symbolisch-intelligente Systeme bauen, und sei es in Form von semantischen Ontologien<sup>19</sup>, die das Internet der Zukunft tragen sollen.<sup>20</sup> Neben der in den 1980er Jahren wichtigen Funktion, auf die Unbrauchbarkeit des symbolischen Ansatzes für die (starke) AI hinzudeuten – wobei Maturanas Biologie der Kognition eine Schlüsselrolle in der Argumentation einnimmt –, sind es wohl zwei Aspekte, die das Buch auch für die Zukunft zu einem Schlüsselwerk machen. Zum einen weist es auf die immanent wichtige Rolle des „Intuitiv-Unbewussten“ für die AI und die Kognitionswissenschaft und damit auf die Frage nach dem Ursprung unserer Wirklichkeitskonstruktionen hin (Riegler 2003). Und zum anderen ist dank des Buches die Bereitschaft gestiegen, sich in diesen Disziplinen mehr mit philosophischen Grundlagen auseinanderzusetzen. Wenn auch die konstruktivistischen Aspekte des Buches bei der allgemeinen Leserschaft das geringste Echo hervorgerufen zu haben scheinen, so kann in Anbetracht der „Verbiologisierung“ der AI doch damit gerechnet werden, dass sie das in Zukunft mit konstruktivistischen Konzepten wie Strukturkoppelung und konstruktivistischen Definitionen wie „Kognition = Leben“ (Maturana/Varela 1980: 13; Stewart 1996) noch tun werden. Terry Winograd und Fernando Flores haben dafür einen wichtigen Grundstein gelegt, indem sie die The-

---

<sup>19</sup> In der Informatik ist eine Ontologie die formale Definition von semantischen Beziehungen zwischen einer Menge von Begriffen und einem Gegenstandsbereich sowie den darauf arbeitenden Inferenzmechanismen.

<sup>20</sup> Dass Computer entgegen früheren Annahmen noch immer nicht in der Lage sind, beliebige visuelle Information zu verstehen, stört heute niemanden mehr – im Gegenteil. Auf dieser Unfähigkeit bauen Sicherheitssysteme auf, die beispielsweise bei der Registrierung auf Internetseiten dem Benutzer die scheinbar unsinnige Frage stellen, von einem Bild die darin enthaltene, verzerrte oder teilweise überdeckte Buchstabenfolge abzutippen, weil man so sicher sein kann, dass es ein menschlicher Benutzer ist und kein Internetroboter.

orie der Autopoiese erfolgreich als Instrument gegen die mechanistisch-realistische Erkenntnistheorie einsetzen, die der AI zugrunde lag. Aufgrund der persönlichen Lebensgeschichte der Autoren blieb das konstruktivistische Element allerdings auf Humberto R. Maturanas und Francisco J. Varelas Konstruktivismus beschränkt. Zwar wird beispielsweise Heinz von Foersters Einfluss auf Maturana ausdrücklich gewürdigt, seine eigenen Ideen zur Kognition und zu einer alternativen Artificial Intelligence, wie sie an seinem Biological Computer Lab erarbeitet wurden (Müller/Müller 2007), fanden aber keine Berücksichtigung. Auch Ernst von Glasersfelds Beiträge zu einem konstruktivistischen Verständnis der in der AI so zentralen Themen wie Lernen und Wissen wären als Bereicherung der Argumentation im Buch vorstellbar. Sind es aber nicht gerade die „unvollendeten“ Werke, die Neugier wecken und den Ansporn zu weiteren Verbesserungen bieten? *Understanding Computers and Cognition* ist eines dieser Schlüsselwerke, die das Potenzial des Konstruktivismus für Anwendungen über die Kernbereiche des Konstruktivismus hinaus (Riegler 2001 a, b) auch in technischen Disziplinen aufweisen.

## Literatur

- Bobrow, Daniel G./Patrick J. Hayes (1985): Artificial Intelligence – Where Are We? In: *Artificial Intelligence*. 25. Jg. H. 3. S. 375–415.
- Boden, Margaret A. (2008): *Mind as Machine. A History of Cognitive Science*. Oxford: Clarendon.
- Brooks, Rodney A. (1991): Intelligence Without Representation. In: *Artificial Intelligence*. 47. Jg. H. 1–3. S. 139–159.
- Cariani, Peter (1993): To Evolve an Ear. Epistemological Implications of Gordon Pask's Electrochemical Devices. In: *Systems Research*. 10. Jg. H. 3. S. 19–33.
- Clancey, William J./Stephen W. Smoliar/Mark J. Stefik (Hrsg.) (1994): *Contemplating Minds. A Forum for Artificial Intelligence*. Cambridge, MA: MIT Press.
- Csikszentmihályi, Mihály (1999): *Das flow-Erlebnis. Jenseits von Angst und Langeweile: im Tun aufgehen*. Stuttgart: Klett-Cotta.
- De Michelis, Giorgio (2008): The Contribution of the Language-Action Perspective to a New Foundation for Design. In: Thomas Erickson/David W. McDonald (Hrsg.): *HCI Remixed. Essays on Works that Have Influenced the HCI Community*. Cambridge, MA: MIT Press. S. 293–297.
- Dennett, Daniel C. (1984): Cognitive Wheels. The Frame Problem of AI. In: Christopher Hookway (Hrsg.): *Minds, Machines, and Evolution*. London: Cambridge Univ. Press.
- Dreyfus, Hubert L. (1965): *Alchemy and Artificial Intelligence*. In: Research Report P-3244. Santa Monica, CA: Rand Corporation. URL: <http://www.rand.org/pubs/papers/2006/P3244.pdf>. (Abgerufen am 29.5.2010)
- Dreyfus, Hubert L. (1972): *What Computers Can't Do. A Critique of Artificial Reason*. San Francisco: W. H. Freeman.
- Dreyfus, Hubert L. (2007): Why Heideggerian AI Failed and How Fixing It Would Require Making It More Heideggerian. In: *Artificial Intelligence*. 171. Jg. H. 18. S. 1137–1160.
- Fisher, Lawrence M. (2009): Fernando Flores Wants to Make You an Offer. In: *strategy+business*. H. 57. (24.11.2009). URL: <http://www.strategy-business.com/article/09406?gko=ce081>. (Abgerufen am 19.6.2010)
- Fogel, Lawrence J./Alvin J. Owens/Michael J. Walsh (1996): *Artificial Intelligence through Simulated Evolution*. New York: Wiley.

- Freksa, Christian (1988): Intrinsische vs. extrinsische Repräsentation zum Aufgabenlösen oder die Verwandlung von Wasser in Wein. In: Gerhard Heyer/Josef Krems/Günther Görz (Hrsg.): Wissensarten und ihre Darstellung. Beiträge aus Philosophie, Psychologie, Informatik und Linguistik. Berlin: Springer. [= Informatik-Fachberichte 169]. S. 155–165.
- Glaserfeld, Ernst von (1997): Radikaler Konstruktivismus. Frankfurt am Main: Suhrkamp.
- Goldberg, David E. (1989): Genetic Algorithms in Search, Optimization and Machine Learning. Boston, MA: Kluwer.
- Harnad, Stevan (1990): The Symbol Grounding Problem. In: *Physica D*. 42. Jg. H. 1–3. S. 335–346.
- Harmelen, Frank van/Vladimir Lifschitz/Bruce Porter (Hrsg.) (2008): Handbook of Knowledge Representation. Amsterdam: Elsevier.
- Koza, John R. (1992): Genetic Programming. On the Programming of Computers by Means of Natural Selection. Cambridge, MA: MIT Press.
- Maturana, Humberto J. (1985): Erkennen. Die Organisation und Verkörperung von Wirklichkeit. 2. durchgesehene Aufl. Braunschweig/Wiesbaden: Vieweg.
- Maturana, Humberto R./Francisco J. Varela (1980): Autopoiesis and Cognition. The Realization of the Living. Boston: Reidel.
- McCarthy, John/Marvin Minsky/Nathan Rochester/Claude Shannon (1955): A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. URL: <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>. (Abgerufen am 19.6.2010)
- Mohnhaupt, Michael/Klaus Rehkämper (1990): Gedanken zu einer neuen Theorie der Kognition. Das Buch von T. Winograd/F. Flores: Erkenntnis – Maschinen – Verstehen. Zur Neugestaltung von Computersystemen. In: *Kognitionswissenschaft*. 1. Jg. H. 3. S. 36–45.
- Müller, Albert/Karl H. Müller (Hrsg.) (2007): An Unfinished Revolution? Heinz von Foerster and the Biological Computer Laboratory | BCL 1958–1976. Wien: edition echoraum.
- Neisser, Ulric (1967): *Cognitive Psychology*. New York: Meredith.
- Neisser, Ulric (1976): *Cognition and Reality*. San Francisco: Freeman.
- Newell, Alan/Herbert A. Simon (1963): GPS – A Program that Simulates Human Thought. In: Edward A. Feigenbaum/Julian Feldman (Hrsg.): *Computers and Thought*. New York: McGraw-Hill. S. 279–293.
- Newell, Alan/Herbert A. Simon (1976): Computer Inquiry as an Empirical Inquiry. Symbols and Search. In: *Communications of the Association for Computing Machinery*. 19. Jg. H. 3. S. 113–126.
- Partridge, Derek (1987): Human Decision Making & the Symbolic Search Space Paradigm in AI. In: *AI & Society*. 1. Jg. H. 2. S. 103–114.
- Polanyi, Michael (1966): *The Tacit Dimension*. Garden City, NY: Doubleday & Co.
- Pörksen, Bernhard (2002): *Die Gewissheit der Ungewissheit. Gespräche zum Konstruktivismus*. Heidelberg: Carl-Auer-Systeme.
- Rechenberg, Ingo (1973): *Evolutionsstrategie. Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Stuttgart: Fromman-Holzboog.
- Riegler, Alexander (Hrsg.) (2001 a): Sonderheft „The Impact of Radical Constructivism on Science. Part 1: The Paradigm, Biology and Cognition.“ *Foundations of Science*. 6. Jg. H. 1–3. URL: <http://www.univie.ac.at/constructivism/pub/fos>. (Abgerufen am 19.6.2010)
- Riegler, Alexander (Hrsg.) (2001 b): Sonderheft „The Impact of Radical Constructivism on Science. Part 2: Physics, Reasoning and Learning.“ *Foundations of Science*. 6. Jg. H. 4. URL: <http://www.univie.ac.at/constructivism/pub/fos>. (Abgerufen am 19.6.2010)
- Riegler, Alexander (2002): When Is a Cognitive System Embodied? In: *Cognitive Systems Research*. [Sonderheft „Situated and Embodied Cognition“.] 3. Jg. H. 3. S. 339–348. URL: <http://www.univie.ac.at/constructivism/people/riegler/pub/?retrieve=24>. (Abgerufen am 19.6.2010)
- Riegler, Alexander (2003): Whose Anticipations? In: Martin V. Butz/Olivier Sigaud/Pierre Gérard (Hrsg.): *Anticipatory Behavior in Adaptive Learning Systems. Foundations, Theories, and Systems*. Lecture Notes in Artificial Intelligence. Dordrecht: Springer. S. 11–22.

- Riegler, Alexander (2007): The Radical Constructivist Dynamics of Cognition. In: Brendan Wallace/Alastair Ross/John Davies/Tony Anderson (Hrsg.): *The Mind, the Body and the World. Psychology after Cognitivism?* London: Imprint. S. 91–115.
- Rubin, Harriet (1998): The Power of Words. In: *Fast Company*. H. 21. URL: <http://www.fastcompany.com/magazine/21/flores.html>. (Abgerufen am 18.6.2010)
- Searle, John R. (1980): Minds, Brains and Programs. In: *Behavioral and Brain Sciences*. 3. Jg. H. 3. S. 417–457.
- Searle, John R. (1997): Was ist Realismus? In: *Die Konstruktion der gesellschaftlichen Wirklichkeit. Zur Ontologie sozialer Tatsachen*. Reinbek bei Hamburg: Rowohlt. S. 162–170.
- Stewart, J. (1996): Cognition = Life: Implications for Higher-Level Cognition. In: *Behavioural Processes*. 35. Jg. S. 311–326
- Turing, Alan (1950): Computing Machinery and Intelligence. In: *Mind*. 59. Jg. H. 236. S. 433–460.
- Van Gelder, Tim (1998): The Dynamical Hypothesis in Cognitive Science. In: *Behavioral and Brain Sciences*. 21. Jg. H. 5. S. 615–628.
- Varela, Francisco J. (1990): *Kognitionswissenschaft – Kognitionstechnik*. Frankfurt am Main: Suhrkamp.
- Varela, Francisco J./Evan Thompson/Eleanor Rosch (1991) *The Embodied Mind. Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.
- Wegner, Daniel M. (2002): *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.
- Weizenbaum, Joseph (1966): ELIZA – A Computer Program for the Study of Natural Language Communication between Man and Machine. In: *Communications of the Association for Computing Machinery*. 9. Jg. H. 1. S. 36–45.
- Winograd, Terry (1972): *Understanding Natural Language*. New York: Academic Press.
- Winograd, Terry (1983): *Language as a Cognitive Process*. Vol. I: Syntax. Reading/MA: Addison-Wesley.
- Winograd, Terry (1990): Thinking Machines. Can There Be? Are We? In: Derek Partridge/Yorick Wilks (Hrsg.): *The Foundations of Artificial Intelligence*. Cambridge: Cambridge Univ. Press. S. 167–189.
- Winograd, Terry/Fernando Flores (1986): *Understanding Computers and Cognition. A New Foundation for Design*. Norwood/NJ: Ablex.
- Winograd, Terry/Fernando Flores (1989): *Erkenntnis Maschinen Verstehen. Zur Neugestaltung von Computersystemen*. Berlin: Rotbuch.
- Winograd, Terry/Fernando Flores (1994): A Response to the Reviews. In: William J. Clancey/Stephen W. Smoliar/Mark J. Stefik (Hrsg.): *Contemplating Minds. A Forum for Artificial Intelligence*. Cambridge/MA: MIT Press. S. 210–221.