

Arbeitsbericht Nr. 32

Schwache Argumente gegen die starke KI-These

**- Ein Beitrag zur Diskussion über
SEARLE's chinesisches Zimmer -**

von

Dr. Stephan Zelewski

Köln 1990

Alle Rechte vorbehalten.

Abstract

Jüngst haben J.R. SEARLE einerseits sowie P.M. CHURCHLAND und P.S. CHURCHLAND andererseits in zwei aufeinander bezogenen Beiträgen den Streit um die starke KI-These wiederbelebt. Es erstaunt, daß diese Grundlagendiskussion immer noch in einer Weise ausgetragen wird, die allgemein anerkannten wissenschaftlichen Argumentationskriterien zuwiderläuft. Daher wird aufgezeigt, daß die starke KI-These so, wie sie die drei Autoren vortragen, nicht ernsthaft aufrechterhalten werden kann. Es wird eine Reformulierung der starken KI-These angeboten, welche diese Schwierigkeiten vermeidet. Allerdings läßt sich SEARLE's Auffassung von der KI-Forschung mit dieser Reformulierung nicht vereinbaren. Wird jedoch an seiner ursprünglichen Variante der starken KI-These festgehalten, so leidet seine angebliche Widerlegung dieser These unter schwerwiegenden Mängeln. Schließlich wird gezeigt, wie die Thesenwiderlegung logisch schlüssig erfolgen kann, dann aber trivial und unbefriedigend bleibt.

Inhaltsverzeichnis

	Seite
1 Exposition	1
2 Die essentialistische Argumentationsfalle	2
3 Eine Reformulierung der starken KI-These	5
4 Ein Fehlschluß	9
5 Eine stringente, aber dennoch unbefriedigende Widerlegung der starken KI-These	12

1 Exposition

Jüngst haben J.R. SEARLE [1] einerseits sowie P.M. CHURCHLAND und P.S. CHURCHLAND [2] andererseits in zwei aufeinander bezogenen Beiträgen den Streit um die starke KI-These wiederbelebt [3]. Es überrascht, daß diese Grundlagendiskussion immer noch in einer Weise ausgetragen wird, die allgemein anerkannten wissenschaftlichen Argumentationskriterien zuwiderläuft. In der Hauptsache leidet die Thematisierung der starken KI-These unter drei Schwächen:

- o Sowohl SEARLE als auch CHURCHLAND/CHURCHLAND stützen ihre Argumentation auf eine essentialistische Begriffsbildung. Eine solche begriffliche Basis findet heute seitens der Wissenschaftstheorie kaum noch einen namhaften Vertreter.
- o Selbst wenn die höchst fragwürdige begriffliche Basis hingenommen wird, so kann doch SEARLE's angebliche Widerlegung der starken KI-These überhaupt nicht überzeugen. Denn er legt eine "Beweisführung" vor, die keineswegs stringent ist.
- o Seine Argumentation läßt sich zwar wohlwollend so rekonstruieren, daß sie logisch zwingend wird. Doch dann ist seine Beweisführung trivial. Der eigentlich interessante Aspekt wird dabei in ein Axiom eingeschlossen und als solches nicht mehr gerechtfertigt.

SEARLE SOWIE CHURCHLAND/CHURCHLAND stimmen in ihren Formulierungen der starken KI-These insoweit überein, als sie gemeinsam die Wesensidentität von menschlichem Denken einerseits und Computerprogrammen andererseits behandeln. Besonders deutlich wird dies bei SEARLE. Ihm zufolge besagt die starke KI-These:

[1] Vgl. SEARLE, J.R.: Ist der menschliche Geist ein Computerprogramm?; in: Spektrum der Wissenschaft, Jg. 1990, Heft 3, S. 40-47; im folgenden zitiert als: SEARLE (1990).

[2] Vgl. CHURCHLAND, P.M.; CHURCHLAND, P.S.: Ist eine denkende Maschine möglich?; in: Spektrum der Wissenschaft, Jg. 1990, Heft 3, S. 47-54; im folgenden zitiert als: CHURCHLAND/CHURCHLAND (1990).

[3] Ausgangspunkt der Kontroverse ist SEARLE's Beitrag zur starken KI-These in seinem Artikel SEARLE, J.R.: Minds, Brains, and Programs; in: Behavioral and Brain Sciences, Vol. 3 (1980), S. 417-458. Er wurde mehrfach wiederaufgenommen, z.B. in: HOFSTADTER, D.R.; DENNETT, D.C. (Hrsg.): The Mind's I - Fantasies and Reflections on Self and Soul, New York 1981, S. 353-373; sowie in: HAUGELAND, J. (Hrsg.): Mind Design - Philosophy, Psychology, Artificial Intelligence, 2. Druck, Cambridge (Massachusetts) - London 1982, S. 282-306.

Ein Computerprogramm, das den Turing-Test besteht, ist "ein denkender Geist im gleichen Sinne, wie es der menschliche Geist ist" [4]. Es werde unterstellt, "daß Computer zwangsläufig dächten, weil, was sie tun, alles sei, was Denken ausmacht" [5]. "Die Behauptung der starken KI lautet, daß jedes ... rechenfähige System ... Gedanken und Gefühle haben müsse, sofern es nur mit dem adäquaten Programm gefüttert wird" [6].

Dabei wird der menschliche Geist mit Begriffen wie Einsicht, Verständnis, geistiger Gehalt, intrinsische Bedeutung, Symbolbedeutung und Semantik assoziiert.

2 Die essentialistische Argumentationsfalle

Sowohl SEARLE als auch CHURCHLAND/CHURCHLAND unterstellen in ihrer gesamten Argumentation, daß ein gemeinsames "Wesen", eine "Essenz" des Begriffs "menschlicher Geist" und aller seiner Assoziationen real existiere [7]. Sie glauben, dieses Begriffswesen erkannt zu haben. Schließlich behaupten sie, das menschliche Denken erfülle tatsächlich diese reale Begriffswesenheit. Dieses wesensmäßige Verständnis des menschlichen Geistes scheint den Autoren so selbstverständlich zu sein, daß sie sich an keiner Stelle verpflichtet sehen, den Begriff klar zu definieren. Stattdessen unterstellen sie stets ein diffuses umgangssprachliches Vorverständnis darüber, was das "Wesen" des menschlichen Geistes, der Semantik und aller weiteren Assoziationen sei.

Ein solcher Begriffsrealismus ist jedoch vollkommen unbefriedigend. Die Unterstellung, es gäbe so etwas wie ein "Wesen des menschlichen Geistes" bedeutet einen Rückfall in aristotelische Philosophietradition. Seitens der Wissenschaftstheorie wird sie heute nahezu unisono abgelehnt [8]. Da sich reale "Begriffswesen-

[4] SEARLE (1990), S. 40.

[5] SEARLE (1990), S. 42.

[6] SEARLE (1990), S. 44.

[7] SEARLE und CHURCHLAND/CHURCHLAND verwenden Formulierungen wie "Wesen eines Begriffs" zwar nicht explizit. Aber ihre gesamte Argumentationsführung folgt der essentialistischen Diktion. An einer Stelle offenbart SEARLE sogar explizit seine Verwurzelung im Begriffsrealismus, wenn er von physikalischen und biologischen Merkmalen spricht, die dem menschlichen Geist angeblich "essentiell" (SEARLE (1990), S. 47) zukommen.

[8] Vgl. z.B. WEINGARTNER, P.: Kann man von Definitionen sagen, daß sie wahr oder falsch sind?; in: Ratio, 7. Bd. (1965), S. 55-82; ESSER, H.; KLENOVITZ, K.; ZEHNPFFENNIG, H.: Wissenschaftstheorie 1: Grundlagen und Analytische Wissenschaftstheorie, Stuttgart 1977, S. 56ff.; PUTNAM, H.: Modell und Wirklichkeit; in: Conceptus, 16. Jg. (1982), Nr. 38, S. 9-30, insb. S. 28ff.; GADENNE, V.: Theorie und Erfahrung in der

heiten" nicht intersubjektiv überprüfen lassen, können sie beliebig postuliert werden, ohne jedoch einer kritisch-rationalen Diskussion zugänglich zu sein. Auch auf sprachphilosophische Arbeiten VON HUMBOLDT'S [9] oder WEISGERBER'S [10] ließe sich verweisen, denen zufolge Begriffsbildungen schöpferische Akte des Menschen - ein "Prozeß des Wortens der Welt" [11] - darstellen, aber keinerlei Erkenntnis über Wirkliches enthalten.

Wer dennoch in essentialistischer Tradition explizit auf das "Wesen" eines Begriffes rekurriert oder die umgangssprachliche Kenntnis eines solchen Begriffswesens implizit voraussetzt, begibt sich in eine schwierige Argumentationsposition: Bevor seine weiterführende Argumentation überhaupt erörtert werden kann, müßte er zunächst einmal nachweisen, daß die von ihm hypostasierten Begriffswesenheiten real existieren und auch erkannt werden können. Da sich dies seit Jahrhunderten immer wieder als erfolgloses und unfruchtbares Unterfangen herausgestellt hat, beziehen die Anhänger des Begriffsrealismus eine Position, die von vornherein zum Scheitern verurteilt ist. Dies ist die essentialistische Argumentationsfalle des Begriffsrealismus. Um so mehr überrascht es, daß sich sowohl SEARLE als auch CHURCHLAND/CHURCHLAND in der essentialistischen Diktion bewegen, ohne auch nur einen Gedanken an die Schwächen des Begriffsrealismus zu verlieren.

Falls SEARLE tatsächlich die starke KI-These angreifen wollte, müßte er zunächst intersubjektiv nachvollziehbar definieren, was er unter dem menschlichen Geist verstehen möchte. Dann müßte er nachweisen, daß sich die vorgelegte Definition auf Menschen tatsächlich zutrifft. (Dies dürfte nicht schwerfallen.) Schließlich müßte er den entscheidenden Nachweis liefern, daß Computerprogramme das - nunmehr wohldefinierte - Leistungsvermögen des menschlichen Geistes niemals erreichen können. Stattdessen operiert SEARLE jedoch fortlaufend mit einem undefinierten Geistes-

psychologischen Forschung, Tübingen 1984, S. 177; QUINE, W.V.Q.: Theorien und Dinge, Frankfurt 1985, S. 149; ALBERT, H.: Kritik der reinen Erkenntnislehre - Das Erkenntnisproblem in realistischer Perspektive, Tübingen 1987, S. 66ff. Eine Ausnahme stellen allenfalls die Anhänger der phänomenologischen Wesensschau HUSSERL'scher Prägung dar. Da sie jedoch für die KI-Forschung keine fruchtbaren Beiträge zu leisten vermochten, werden sie hier nicht weiter berücksichtigt.

[9] Vgl. VON HUMBOLDT, W.: Werke in fünf Bänden, Bd. III: Schriften zur Sprachphilosophie, Darmstadt 1963, S. 386ff. u. 426ff., insb. S. 386, 426, 430, 435f. u. 438f.

[10] Vgl. WEISGERBER, L.: Die geistige Seite der Sprache und ihre Erforschung, Düsseldorf 1971, S. 149ff., insb. S. 149 u. 153ff.; WEISGERBER, L.: Die anthropologische Tragweite der energetischen Sprachbetrachtung; in: GADAMER, H.-G.; VOGLER, P. (Hrsg.): Neue Anthropologie, Bd. 7: Philosophische Anthropologie, Zweiter Teil, Stuttgart 1975, S. 168-203, darin S. 176ff u. 189ff.

[11] WEISGERBER (1971), S. 155; ähnlich auch WEISGERBER (1975), S. 174.

und Semantikbegriff. Er appelliert lediglich an eine diffuse umgangssprachliche Intuition, was das "Wesen" dieser Begriffe sei. Dabei beruft er sich auf "eine selbstevidente logische Grundwahrheit" [12], die unmittelbar erkennen lasse, worin der wesensmäßige Unterschied zwischen menschlichem Geist und Computerprogrammen liege. Ebenso postuliert er, "daß dem menschlichen Geist irgend etwas essentiell Physikalisches und Biologisches anhaftet" [13], das Computerprogrammen als solchen - also als reiner Software ohne Berücksichtigung ihrer Implementierung in einer realen Ablaufumgebung - prinzipiell nicht zukommen könne. Bezugnahmen auf "selbstevidente" Einsichten oder "irgend etwas" können aber nicht als seriöses Argumentationsfundament akzeptiert werden. Vielmehr offenbaren sie die inhaltliche Leere von SEARLE'S Begriffsrealismus, der zwar das "Wesen" des menschlichen Geists immer wieder voraussetzt, es aber nirgends konkret zu erläutern vermag.

Dies verrät auch eine unseriöse Argumentationsführung: SEARLE benutzt einen undefinierten Geistesbegriff. Vermittels des diffusen umgangssprachlichen Begriffsverständnisses transportiert er die unterschwellige "Botschaft", daß solches geistige Vermögen doch nur dem Menschen, aber nicht den "geistlosen" Computern zukommen könne. Hierbei wird aber nicht mehr wissenschaftlich argumentiert, sondern nur noch alltagsweltlich suggeriert.

Schließlich ist SEARLE vorzuhalten, daß er einen typischen Imputationsbegriff verwendet. Das, was er unter dem menschlichen Geist zu verstehen scheint, läßt sich nicht anhand intersubjektiv überprüfbarer Kriterien definieren, sondern einem Erkenntnisobjekt nur per Setzung zuschreiben oder aberkennen [14]. Dann fällt es SEARLE natürlich nicht schwer, dem menschlichen Denken die Attribute "vernünftig" oder "semantisch" zuzuordnen, sie jedoch Computerprogrammen abzusprechen. Da er keinen wohldefinierten Geistesbegriff verwendet, lassen sich diese impliziten

[12] SEARLE (1990), S. 45.

[13] SEARLE (1990), S. 47.

[14] Darin gleicht der Begriff "Geist" anderen Imputationsbegriffen wie der Kausalität oder dem Regelfolgen. Denn auch bei realen Prozessen, die in der Anschauungsform "Zeit" ablaufen, läßt sich ein kausales Gesetz, das den Ereigniseintritten zugrundeliegen soll, strenggenommen nicht rein empirisch feststellen. Es kann nur durch Setzung des erkennenden Subjekts dem beobachteten Prozeß zugeschrieben werden. Dies ist die alt bekannte, an HUME anschließende Problematik des Schlusses vom "post hoc" auf das "propter hoc". Ebenso läßt sich aus dem Beobachten eines endlichen menschlichen Verhaltens, das regulär erscheint, nicht mit Sicherheit eine bestimmte Regel ableiten, die diesem Verhalten zugrundeliegt. Das Befolgen einer bestimmten Regel kann einer endlichen Verhaltenssequenz nur zugeschrieben werden. Das haben in jüngerer Zeit vor allem KRIPKE und STEGMÜLLER herausgearbeitet, die sich auf WITTGENSTEIN'S Erörterung des Regelfolgens bezogen.

Setzungen im diffusen umgangssprachlichen Begriffsverständnis vortrefflich verbergen. Versuchte SEARLE hingegen, seinen Geistesbegriff nachvollziehbar zu definieren, so würde sich schnell zeigen, daß er in eine von zwei Sackgassen geraten würde, falls er am Geist als "wesentlich" menschlicher Qualität festhalten möchte:

- o Entweder müßte er eine *petitio principii* begehen, indem er in der Geistesdefinition bereits auf den Menschen als Geistes-träger Bezug nähme. Dann wäre die Widerlegung der starken KI-These unproblematisch, aber trivial, weil die semantische Qualität per definitionem für den Menschen reserviert wäre.
- o Oder es würde sich zeigen, daß der Geistesbegriff nicht ausschließlich anhand intersubjektiv überprüfbarer Kriterien definiert werden kann. Wollte SEARLE dann immer noch gegen die starke KI-These opponieren, so müßte er wieder zu subjektiven Setzungen wie den oben kritisierten greifen. Das wäre aber keine überzeugende argumentative Widerlegung, sondern nichts weiter als der definatorische Trick eines Imputationsbegriffs.

Angesichts dieser beiden Alternativen verwundert es nicht, daß SEARLE eine klare Definition dessen, was er als menschlichen Geist oder Semantik versteht, konsequent vermieden hat. Darüber hinaus scheint ihm entgangen zu sein, daß alle Programmiersprachen (mindestens) eine wohldefinierte Semantik besitzen, allerdings eine formale Semantik. Schon dieses Faktum hätte Anlaß genug sein müssen, sein eigenes, abweichendes Semantikverständnis präzise darzulegen.

3 Eine Reformulierung der starken KI-These

Wegen der voranstehend erörterten Probleme essentialistischer Begriffsbildung wird die starke KI-These so, wie sie eingangs in Anlehnung an SEARLE und CHURCHLAND/CHURCHLAND skizziert wurde, heute von keinem seriösen Vertreter der KI-Forschung ernsthaft vertreten.

Es trifft zwar zu, daß sie in den frühen Phasen der Erforschung Künstlicher Intelligenz artikuliert wurde. Doch handelte es sich damals um Versuche einer noch jungen, zumeist belächelten Wissenschaftsrichtung, sich durch markante Postulate Gehör und Anerkennung zu verschaffen. Solche Äußerungen sollten daher aus wissenschaftssoziologischer Perspektive wohlwollend zur Kenntnis

genommen, aber aus wissenschaftstheoretischer Perspektive nicht im Sinne der o.a. starken KI-These überzeichnet werden. Wer heute dennoch gegen solche Übertreibungen von historischen Varianten der starken KI-These opponiert, der scheint nicht an einer ernsthaften Auseinandersetzung mit den Grundlagen der Künstlichen Intelligenz interessiert zu sein. Vielmehr scheint er einen Popanz aufbauen zu wollen, der sich nachträglich um so leichter "widerlegen" läßt.

Stattdessen sollte heute die starke KI-These deutlich vorsichtiger formuliert werden, um der essentialistischen Begriffsfalle zu entgehen. Beispielsweise ließe sich als revidierte starke KI-These aufstellen [15]:

Der menschliche Geist läßt sich in dem Ausmaß, in dem sein Leistungsvermögen durch sprachlich wohldefinierte und intersubjektiv nachprüfbarere Voraussetzungen und Resultate von Denkprozessen beschrieben wird, durch leistungsäquivalente automatische Informationsverarbeitungssysteme (Computerprogramme) rekonstruieren.

Dabei wird auf keinerlei "Wesen" von Geist oder Semantik Bezug genommen. A fortiori wird auch keine Wesensidentität von Menschen und Computern behauptet. Stattdessen verfolgt die derart reformulierte starke KI-These ein beachtenswertes Erkenntnisprogramm [16]: Indem sie an die Wohldefiniertheit und Nachprüfbarkeit von Denkprozessen anknüpft, hat sie dazu geführt, daß auf etlichen Feldern erst durch die KI-Forschung Bemühungen eingeleitet wurden, kognitive Leistungen des Menschen aus dem Bereich vager umgangssprachlicher Andeutungen herauszulösen, um sie einer präzisen und nachprüfbareren Begrifflichkeit zuzuführen. Dieser befruchtende KI-Beitrag zur Kognitionsforschung sollte nicht unterschätzt werden. Zugleich läßt sich vermuten, warum gerade aus geisteswissenschaftlicher Richtung mitunter so vehement gegen anachronistische Varianten der starken KI-These opponiert wird. Der Nachdruck, mit dem die KI-Forschung auf operationaler

[15] Es ist eine Frage definitorischer Setzung, ob diese vorsichtige Reformulierung noch als Variante der starken KI-These angesehen oder schon der schwachen KI-These zugerechnet wird. SEARLE (1990), S. 40, spricht von der schwachen KI-These nur in dem instrumentellen Sinne, daß Computerprogramme "brauchbare Hilfsmittel beim Studium des menschlichen Geistes" darstellen. Bei der oben vorgelegten Reformulierung werden aber informationsverarbeitende Automaten nicht bloß als Instrumente zur Erforschung des menschlichen Geistes, sondern als dessen Rekonstruktionen betrachtet. Daher wird sie weiterhin der starken KI-These zugerechnet.

[16] Vgl. MINSKY, M.: Why People Think Computers Can't; in: The AI Magazine, Vol. 3 (1982), No. 4, S. 3-15, insb. S. 12f. u. 14f.

Begriffsbildung beharrt, scheint manchen Geisteswissenschaftlern zu mißhagen.

Die oben angeführte vorsichtige Reformulierung der starken KI-These unterscheidet sich von der starken KI-These, wie sie von SEARLE und CHURCHLAND/CHURCHLAND vertreten wird, vor allem dadurch, daß das diffuse umgangssprachliche Vorverständnis vom "Wesen" des menschlichen Geists durch "sprachlich wohldefinierte und intersubjektiv nachprüfbarere Voraussetzungen und Resultate von Denkprozessen" ersetzt wird. Auf diese Weise wird der oftmals geschmähte TURING-Test wieder aufgewertet: In dem Ausmaß, in dem das Leistungsvermögen des menschlichen Denkens durch präzise Spezifizierung seiner Voraussetzungen und Resultate beschrieben wird, kann ein entsprechender TURING-Test formuliert werden. Die reformulierte starke KI-These postuliert dann, daß es grundsätzlich immer möglich ist, für jeden derart wohldefinierten TURING-Test einen informationsverarbeitenden Automaten zu konstruieren, der diesen Test zu bestehen vermag. Dabei braucht der TURING-Test keineswegs auf seine historische, einseitig auf sprachliches Leistungsvermögen konzentrierte Variante beschränkt zu bleiben. Vielmehr haben CHURCHLAND/CHURCHLAND durch ihr instruktives Roboterbeispiel deutlich aufgezeigt [17], daß in einen TURING-Test auch die Voraussetzungen und Resultate eines Denkprozesses eingehen können, die sensorische Umweltwahrnehmungen bzw. effektorsische Umweltbeeinflussungen darstellen.

Die Reformulierung der starken KI-These schließt nicht aus, daß das "Wesen" des menschlichen Geistes in einen Bereich zurückgezogen wird, in dem es unmöglich sein soll, sprachlich wohldefinierte und intersubjektiv nachprüfbarere Voraussetzungen sowie Resultate von Denkprozessen auszudrücken. Diese Argumentationsstrategie liegt vor allem der tacit knowledge-These zugrunde. Allerdings leidet dieser Ansatz unter der Fragwürdigkeit seiner Unmöglichkeitsbehauptung. Denn für einen Aspekt des menschlichen Geistes, der sich jeder präzisen Beschreibung entziehen soll, wird schwerlich eine stringente Argumentation vorgelegt werden können, die rechtfertigt, warum sich dieser Aspekt einer solchen Beschreibung grundsätzlich verschließt. Daher handelt es sich um eine typische Argumentationsstrategie der Immunisierung gegenüber jeder potentiellen Kritik. Solche selbstimmunisierenden Argumentationen lassen sich zwar per constructionem nicht widerlegen,

[17] Vgl. CHURCHLAND/CHURCHLAND (1990), S. 48.

werden aber auch gerade deswegen aus dem Kanon wissenschaftlich anerkannter Argumentationsformen ausgegrenzt.

Hier schließt sich der Kreis zum oben angesprochenen Erkenntnisprogramm der KI-Forschung: Indem sie auf sprachlich wohldefinierten und intersubjektiv nachprüfbaren Formulierungen insiziert, steht sie in erklärtem Gegensatz zu selbstimmunisierenden Konzepten für das "Wesen" des menschlichen Geistes. Sie kann zwar keineswegs nachweisen, daß sich alles, was zum menschlichen Geist gehöre, präzise beschreiben lasse. Aber sie versucht diesen Nachweis auch überhaupt nicht, solange im Unklaren bleibt, was denn diesen menschlichen Geist überhaupt als solchen konkret auszeichne.

Stattdessen ruft das Erkenntnisprogramm der KI-Forschung dazu auf, *zunächst* dasjenige Leistungsvermögen des menschlichen Geistes sprachlich zu präzisieren, über dessen automatengestützte Rekonstruktionsmöglichkeit gestritten werden soll. *Erst dann* setzt die vorsichtig reformulierte starke KI-These mit ihrem Anspruch ein, daß sich dieses wohldefinierte und überprüfbare Leistungsvermögen tatsächlich durch informationsverarbeitende Automaten realisieren läßt.

Diese Zweistufigkeit, zuerst das interessierende kognitive Leistungsvermögen zu präzisieren und es dann erst durch leistungsäquivalente Automaten zu rekonstruieren [18], übersehen aber sowohl SEARLE als auch CHURCHLAND/CHURCHLAND in ihrer eingangs vorgestellten Variante der starken KI-These. Sie überspringen die wichtige erste Präzisierungsstufe, indem sie die starke KI-These so formulieren, daß die "Wesensidentität" von menschlichem Geist und Computerprogrammen behauptet wird. Gründlicher kann das Erkenntnisprogramm der KI-Forschung wohl kaum mißverstanden werden, dem es eben auch darum geht, zur inhaltlichen Präzisierung dessen, was Denken bedeuten kann, beizutragen.

[18] Diese Zweistufigkeit läßt sich aus der Perspektive des TURING-Tests auch so ausdrücken: Das Erkenntnisprogramm der KI-Forschung befaßt sich zunächst damit, das jeweils interessierende kognitive Leistungsvermögen so weit zu präzisieren, daß sich die Erfüllung des Leistungsvermögens anhand eines entsprechend ausgelegten TURING-Tests überprüfen läßt. Danach bemüht sie sich, den Anspruch der starken, aber vorsichtig reformulierten KI-These dadurch einzulösen, daß sie Computerprogramme entwirft, welche den leistungsspezifischen TURING-Test tatsächlich erfüllen.

4 Ein Fehlschluß

Trotz aller begrifflichen Vorbehalte sei akzeptiert, es läge eine klare Definition dessen vor, was SEARLE als menschlichen Geist oder Semantik verstehen möchte. Dann bleibt aber seine Beweisführung, in der er aus seinen drei Axiomen seine erste Folgerung zieht, völlig unzureichend. Denn die behauptete Folgerung läßt sich aus den Axiomen so, wie SEARLE sie formuliert hat, überhaupt nicht deduzieren. Seine Argumentation besteht aus [19]:

- o dem Axiom 1: "Computerprogramme sind formal (syntaktisch)."
- o dem Axiom 2: "Dem menschlichen Denken liegen geistige Inhalte (Semantik) zugrunde."
- o dem Axiom 3: "Syntax an sich ist weder konstitutiv noch hinreichend für Semantik."
- o der Folgerung 1: "Programme sind weder konstitutiv noch hinreichend für Geist."

Kritisch ist dabei vor allem das dritte Axiom [20]. Es schließt keineswegs aus, daß mindestens ein syntaktisches Artefakt existieren kann, das eine - wie auch immer definierte - semantische Qualität besitzt. Zwischen der Existenzmöglichkeit eines syntaktischen und zugleich semantischen Konstrukts und SEARLE's dritten Axiom besteht überhaupt kein logischer Widerspruch. Denn dieses Axiom besagt nur: Syntax "an sich" ist für Semantik nicht "hinreichend" [21], erzwingt also nicht *generell* eine semantische Qualität. Das bedeutet jedoch nicht, daß *einzelne* syntaktische Konstrukte existieren könnten, auf die das Prädikat "semantisch" zutrifft.

Genau darauf haben CHURCHLAND/CHURCHLAND in ihrem treffenden Beispiel vom erleuchteten Zimmer [22] hingewiesen. Sie halten SEARLE darin zu Recht vor, ein *einzelnes* Beispiel - wie sein chinesisches Zimmer - reiche logisch nicht aus, um für *alle* syntaktischen Konstrukte nachzuweisen, sie besäßen kein semantisches

[19] Vgl. SEARLE (1990), S. 41f.

[20] Das wird bereits von CHURCHLAND/CHURCHLAND (1990), S. 50, angedeutet.

[21] Was dabei das "konstitutiv" in der Formulierung des dritten Axioms bedeuten mag, bleibt unverständlich. Falls es das logische "notwendig" vertritt, ist es für die Beweisführung überflüssig. Darüber hinaus wäre schwerlich nachzuvollziehen, wie eine semantische Qualität ohne jede syntaktische Grundlage zustanden kommen soll. Falls ein aliud gemeint ist, läßt sich nicht erkennen, was es jenseits einer hinreichenden oder notwendigen Bedingung wohl bedeuten könnte.

[22] Vgl. CHURCHLAND/CHURCHLAND (1990), S. 50f.

Vermögen. In diesem Vorwurf einer logisch unzulässigen Generalisierung liegt die Bedeutung ihres Gegenbeispiels. Dagegen ist die von SEARLE polemisch attackierte Analogie vollkommen unerheblich.

Wenn SEARLE's chinesisches Zimmer tatsächlich ein stringentes Argument darstellen sollte, müßte er nachweisen, daß sich die dort vorgetragene Argumentation auf *alle* überhaupt denkbaren syntaktischen Konstrukte übertragen läßt. Einen solchen Generalisierbarkeitsnachweis unterläßt SEARLE jedoch. Er definiert noch nicht einmal präzise, was er als ein syntaktisches Konstrukt überhaupt zulassen möchte. Stellen z.B. für ihn die oben erwähnte formale Semantik von Programmiersprachen oder die von CHURCHLAND/CHURCHLAND geistreich ins Spiel gebrachte denotationale Semantik eines Roboters, der reale Objekte in seiner Umgebung über Sensoren wahrzunehmen und mit formalen Objekten in seinem Programm zu identifizieren vermag, syntaktische Konstrukte dar? Der Leser mag darüber ins Grübeln geraten - SEARLE selbst geht hierauf nirgends ein, obgleich er doch den Beitrag von CHURCHLAND/CHURCHLAND studiert hat.

In einer semi-formalen Darstellungsweise, die sich an den prädikatenlogischen Kalkül 1. Stufe anlehnt, läßt sich SEARLE's Fehlschluß verdeutlichen. Dabei wird - wie in der prädikatenlogischen Programmiersprache PROLOG - der Name der einzigen Individuenvariable "X" durch einen Großbuchstaben notiert, während alle Namen der Prädikatskonstanten mit einem Kleinbuchstaben beginnen [23]. Dann gilt für SEARLE's drei Axiome und die dadurch angeblich beweisbare Folgerung:

[23] Individuenkonstanten werden nicht verwendet.

Axiom 1:

$$\Lambda (X): \text{computerprogramm}(X) \rightarrow \text{syntaktisch}(X)$$

Axiom 2:

$$\Lambda (X): \text{menschliches_denken}(X) \rightarrow \text{semantisch}(X)$$

Axiom 3:

$$\neg(\Lambda (X): \text{syntaktisch}(X) \rightarrow \text{semantisch}(X))$$

Folgerung 1:

$$\Lambda (X): \text{computerprogramm}(X) \rightarrow (\neg(\text{menschliches_denken}(X)))$$

Es existiert kein logisch zulässiger Schluß, der die Folgerung 1 aus den Axiomen 1 bis 3 ableitet. Insbesondere läßt sich das dritte Axiom logisch konsistent vereinbaren mit der Existenzaussage:

$$\forall (X): \text{syntaktisch}(X) \wedge \text{semantisch}(X)$$

Ebensowenig widerspricht es dem ersten Axiom, daß es sich bei der Entität aus dieser Existenzaussage um ein Computerprogramm handelt:

$$\forall (X): \text{computerprogramm}(X) \wedge \text{syntaktisch}(X) \wedge \text{semantisch}(X)$$

Die Folgerung 1, daß ein Computerprogramm niemals von der Art des menschlichen Denkens sei, beruht auf der Kontraposition des zweiten Axioms:

$$\Lambda (X): \neg(\text{semantisch}(X)) \rightarrow \neg(\text{menschliches_denken}(X))$$

Da nunmehr ein Computerprogramm nicht mehr ausgeschlossen werden kann, das neben seiner syntaktischen auch eine semantische Qualität besitzt, ist es möglich, daß mindestens ein Computerprogramm existiert, welches die Prämisse der voranstehenden Kontraposition verletzt. Folglich läßt sich die Behauptung, durch die Axiome 1 bis 3 sei "bewiesen", daß ein Computerprogramm unmöglich von der Art des menschlichen Denkens sein könne, nicht aufrechterhalten; q.e.d.

5 Eine stringente, aber dennoch unbefriedigende Widerlegung der starken KI-These

Um SEARLE's "Widerlegung" der starken KI-These zu retten, müßte sein drittes Axiom wesentlich schärfer formuliert werden: "Für alle Entitäten gilt: Wenn sie syntaktisch definiert sind und keine darüber hinausgehende Eigenschaft besitzen, dann können sie keine semantische Qualität aufweisen." Oder äquivalent: "Für alle Entitäten gilt: Wenn sie semantische Qualität aufweisen, dann ist es unmöglich, daß sie syntaktisch definiert sind und keine darüber hinausgehende Eigenschaft besitzen." Darüber hinaus müßte das erste Axiom SEARLE's präzisiert werden zu: "Für alle Entitäten gilt: Wenn sie ein Computerprogramm darstellen, dann sind sie syntaktisch definiert und besitzen keine darüber hinausgehende Eigenschaft."

Unter diesen Voraussetzungen ist es trivial zu beweisen, daß sich Computerprogramme durch keine semantische Qualität auszeichnen können. Daraus folgt dann mit Hilfe von SEARLE's zweitem Axiom, daß ein Computerprogramm niemals von der Art des menschlichen Denkens sein kann. Dies läßt sich wiederum in einer semi-formalen Darstellungsweise verdeutlichen, die nunmehr jedoch zur Prädikatenlogik 2. Stufe gehört [24]:

[24] Denn die Quantoren erstrecken sich nunmehr zum Teil auch auf die Prädikatsvariable "Eigenschaft". Sie steht für jede beliebige Eigenschaft, die einer Entität zukommen kann.

Axiom 1 (revidiert):

$$\begin{aligned} \wedge (X): & \text{ computerprogramm}(X) \\ & \rightarrow (\text{syntaktisch}(X) \\ & \quad \wedge (\wedge (\text{Eigenschaft}): \text{Eigenschaft} \neq \text{syntaktisch} \\ & \quad \quad \rightarrow \neg(\text{Eigenschaft}(X)))) \end{aligned}$$

Axiom 2 (unverändert):

$$\wedge (X): \text{ menschliches_denken}(X) \rightarrow \text{semantisch}(X)$$

Axiom 3 (revidiert):

$$\begin{aligned} \wedge (X): & (\text{syntaktisch}(X) \\ & \quad \wedge (\wedge (\text{Eigenschaft}): \text{Eigenschaft} \neq \text{syntaktisch} \\ & \quad \quad \rightarrow \neg(\text{Eigenschaft}(X)))) \\ & \rightarrow (\neg(\text{semantisch}(X))) \end{aligned}$$

Folgerung 1 (unverändert):

$$\wedge (X): \text{ computerprogramm}(X) \rightarrow (\neg(\text{menschliches_denken}(X)))$$

Zum Beweis der Folgerung 1 werden Axiom 1 und Axiom 3 transitiv verkettet. Daraus folgt:

$$\wedge (X): \text{ computerprogramm}(X) \rightarrow (\neg(\text{semantisch}(X)))$$

Wenn dieses Zwischenergebnis mit der bereits o.a. Kontraposition des Axioms 2:

$$\wedge (X): \neg(\text{semantisch}(X)) \rightarrow \neg(\text{menschliches_denken}(X))$$

wiederum transitiv verknüpft wird, resultiert tatsächlich die zu beweisende Folgerung 1:

$$\wedge (X): \text{ computerprogramm}(X) \rightarrow (\neg(\text{menschliches_denken}(X)))$$

Von Interesse ist aber nicht dieser korrekte Beweis, sondern die Prämisse des dritten Axioms, daß *alle rein syntaktisch definier-*

ten Entitäten [25] unmöglich einen semantischen Charakter besitzen könnten. Für eine Rechtfertigung dieser Prämisse reicht SEARLE'S Beispiel des chinesischen Zimmers nicht aus. Denn erstens fehlt ihm - wie oben ausgeführt - der kritische Nachweis seiner generellen Geltung für alle denkmöglichen rein syntaktischen Konstrukte. Zweitens müßte ein solcher Nachweis, um kritisch überprüft werden zu können, auf einem wohldefinierten Begriff des Semantischen - also dessen, was den menschlichen Geist auszeichnen soll, - beruhen. Eben diese Definition bleibt SEARLE aber schuldig.

Die Unzulänglichkeit des dritten, revidierten Axioms wird besonders deutlich, wenn es in der äquivalenten Gestalt seiner Kontraposition betrachtet wird:

Axiom 3 (revidiert, äquivalent umgeformt):

$$\begin{aligned} \Lambda (X): & \quad \text{semantisch}(X) \\ & \quad \rightarrow (\neg(\text{syntaktisch}(X))) \\ & \quad \vee (\vee (\text{Eigenschaft}): \text{Eigenschaft} \neq \text{syntaktisch} \\ & \quad \quad \quad \wedge \text{Eigenschaft}(X)) \end{aligned}$$

Damit diese Variante des Axioms 3 durch eine Entität, welche die Variable "X" vertritt, erfüllt wird, muß es sich also um eine Entität handeln, die überhaupt nicht syntaktisch definiert ist oder die neben ihrer syntaktischen Definition noch mindestens eine nicht-syntaktische Eigenschaft besitzt. Der erste Fall interessiert nicht weiter, weil alle bislang ernsthaft diskutierten semantischen Konzepte jeweils auf einer syntaktischen Basis beruhen.

Der zweite Fall ist aber in dreierlei Hinsicht aufschlußreich: Erstens müßte zur Rechtfertigung des dritten, revidierten Axioms aufgezeigt werden, daß die semantische Qualität einer Entität tatsächlich impliziert, diese Entität müsse mindestens eine nicht-syntaktische Eigenschaft besitzen. Eine solche Implikation ist jedoch bislang nirgends streng nachgewiesen worden. Zweitens wäre eine solche nicht-syntaktische Eigenschaft als "Essenz" des Semantischen inhaltlich konkret zu definieren, um der oben dargelegten Kritik an undefinierten Begriffsverwendungen zu begegnen.

[25] Eine Entität heißt hier rein syntaktisch definiert, wenn sie syntaktisch definiert ist und durch keine weitere, nicht-syntaktische Eigenschaft ausgezeichnet wird.

Die konkrete Bestimmung einer solchen nicht-syntaktischen Eigenschaft, die eine Entität erst als "semantisch" auszeichnen soll, findet sich in SEARLE'S Ausführungen aber an keiner Stelle. Stattdessen greift er zu nebulösen, inhaltsleeren Formulierungen, die bereits an früherer Stelle kritisiert wurden: SEARLE betrachtet die nicht-syntaktische Eigenart des menschlichen Geistes als "eine selbstevidente logische Grundwahrheit" [26], der zufolge "dem menschlichen Geist irgend etwas essentiell Physikalisches und Biologisches anhaftet" [27]. Drittens müßte für die konkret bestimmte nicht-syntaktische Eigenschaft nachgewiesen werden, daß es sich tatsächlich um die mindestens eine nicht-syntaktische Eigenschaft handelt, sie im Sinne der zuerst angesprochenen Implikation für den semantischen Charakter der Entität *notwendig* ist. Da SEARLE eine solche Eigenschaft noch nicht einmal konkretisiert hat, bleibt er erst recht den Nachweis ihrer Notwendigkeit schuldig.

Folglich läßt sich SEARLE'S Widerlegung der starken KI-These zwar so reformulieren, daß sie logisch schlüssig ist. Dann bleibt sie aber dennoch unbefriedigend, weil ihr drittes Axiom das eigentlich Interessante - die behauptete wesensmäßige Verschiedenheit von rein syntaktisch definierten und semantischen Entitäten - bereits voraussetzt. Auf diese Weise wird aber nicht wissenschaftlich argumentiert, sondern nur axiomatisch postuliert.

[26] Vgl. SEARLE (1990), S. 45.

[27] Vgl. SEARLE (1990), S. 47.

Verzeichnis der Arbeitsberichte des
Seminars für Allgemeine Betriebswirtschaftslehre,
Industriebetriebslehre und Produktionswirtschaft
der Universität zu Köln

(bis Sommer 1986: Seminar für Allgemeine
Betriebswirtschaftslehre und Fertigungswirtschaft)

- Nr. 1: ZELEWSKI,STEPHAN: Entscheidungsmodelle zur Verschrottung von Fertigungshilfsmitteln, Köln 1984.
- Nr. 2: KERN,WERNER; ZELEWSKI,STEPHAN: Ein Zuordnungsmodell für Meßgeräte in Energie-Informations-Systemen, Köln 1985.
- Nr. 3: KERN,WERNER; PETERS,ULRICH: Energiebewirtschaftung in industriellen Betrieben - Bericht über eine Befragung, Köln 1985.
- Nr. 4: BOOS,JOCHEN: Lokalisierung von Meßstellen für ein Informations-System zur Energiebewirtschaftung in industriellen Betrieben - Entwicklung eines OR-Modells mit einem Lösungsvorschlag, Köln 1986.
- Nr. 5: ZELEWSKI,STEPHAN: Ansätze der Künstlichen Intelligenz-Forschung zur Unterstützung der Netzplantechnik, Köln 1986.
- Nr. 6: ZELEWSKI,STEPHAN: Schnittstellen bei betrieblichen Informationssystemen - eine Darstellung aus systemtheoretischer und betriebswirtschaftlicher Sicht, Köln 1986.
- Nr. 7: ZELEWSKI,STEPHAN: Konzepte für Frühwarnsysteme und Möglichkeiten zu ihrer Fortentwicklung durch Beiträge der Künstlichen Intelligenz, Köln 1986.
- Nr. 8: ZELEWSKI,STEPHAN: Das Konzept der unscharfen Mengen unter besonderer Berücksichtigung ihrer linguistischen Interpretation - eine Lösung für unscharfe Probleme?, Köln 1986.
- Nr. 9: ZELEWSKI,STEPHAN: Der tau-Wert: Aspekte eines neueren spieltheoretischen Ansatzes zur fairen Preisbildung aus kostenrechnerischer Perspektive, Köln 1986.
- Nr. 10: ZELEWSKI,STEPHAN: Competitive Bidding aus der Sicht des Ausschreibers - ein spieltheoretischer Ansatz, Köln 1986.
- Nr. 11: ZELEWSKI,STEPHAN: Netztheoretische Ansätze zur Konstruktion und Auswertung von logisch fundierten Problembeschreibungen, Köln 1986.
- Nr. 12: ZELEWSKI,STEPHAN: Netztheoretische Fundierung von parallelen Algorithmen für die Lösung linear-ganzzahliger OR-Modelle, Köln 1986.
- Nr. 13: ZELEWSKI,STEPHAN: Intelligente Informationsbanksysteme - benutzerfreundliche Instrumente für die Informationsvermittlung?, Köln 1986.
- Nr. 14: ZELEWSKI,STEPHAN: Komplexitätstheorie - ihr Beitrag zur Klassifizierung und Beurteilung von Problemen des Operations Research, Köln 1986.
- Nr. 15: ZELEWSKI,STEPHAN: Der Informationsbroker, Köln 1986.
- Nr. 16: ZELEWSKI,STEPHAN: Soziale Verantwortbarkeit von Technologien, Köln 1986.

- Nr. 17: ZELEWSKI,STEPHAN: Expertensysteme - Übersicht über Konzeptionen und betriebswirtschaftliche Anwendungsmöglichkeiten, Köln 1986.
- Nr. 18: ZELEWSKI,STEPHAN: Das Leistungspotential der Künstlichen Intelligenz für Industrieanwendungen - Ein Überblick, Köln 1987.
- Nr. 19: ZELEWSKI,STEPHAN: Expertensysteme im "Büro der Zukunft" - Ein Überblick über Anwendungsperspektiven und Bewertungsaspekte, Köln 1987.
- Nr. 20: KUMMER,SEBASTIAN: Computerunterstützung schöpferischer Forschungs- und Entwicklungsaktivitäten, Köln 1987.
- Nr. 21: ZELEWSKI,STEPHAN: Betriebswirtschaftliche Aspekte des industriellen Einsatzes von Expertensystemen - Anwendungsmöglichkeiten und Bewertung, Köln 1988.
- Nr. 22: ZELEWSKI,STEPHAN: Expertensysteme für Prozeßplanung und -steuerung in der Fabrik der Zukunft - Ein Überblick über Konzepte und erste Prototypen, Köln 1988.
- Nr. 23: ZELEWSKI,STEPHAN: Expertensysteme zur Sicherung der Betriebsbereitschaft in der Fabrik der Zukunft, Köln 1988.
- Nr. 24: ZELEWSKI,STEPHAN: Ansätze zur Bewertung des Einsatzes Künstlicher Intelligenz in Industrieunternehmen - aus produktiver und sozialer Sicht, Köln 1988.
- Nr. 25: HÖLSCHER,ANDREAS: Unterstützung der Forschung und Entwicklung in der Pharmaindustrie durch externe Informationen - Möglichkeiten und Grenzen, Köln 1988.
- Nr. 26: SCHRÖDER,HANS-HORST: Entwicklungsstand und -tendenzen bei PPS-Systemen, Köln 1989.
- Nr. 27: ZELEWSKI,STEPHAN: Eine Metakritik an der Kritik konventioneller Rationalitätsauffassungen durch kulturwissenschaftlich fundierte Konzepte praktischer und prozeduraler Rationalität, Köln 1989.
- Nr. 28: ZELEWSKI,STEPHAN: Petrinetze für die Konstruktion und Konsistenzanalyse von logisch orientierten Problembeschreibungen, Köln 1989.
- Nr. 29: ELLE,HANS-DIETER: Aufgeklärter Konstruktivismus - Ein Beitrag zur erkenntnis- und wissenschaftsprogramatischen Diskussion innerhalb der Betriebswirtschaftslehre, Köln 1989.
- Nr. 30: ZELEWSKI,STEPHAN: CAP-Expertensysteme - Anwendungsaspekte Künstlicher Intelligenz im Bereich der Arbeitsplanung, Köln 1989.
- Nr. 31: DOCHNAL,HANS-GUIDO: Darstellung und Analyse von OPT (Optimized Production Technology) als Produktionsplanungs- und -steuerungskonzept, Köln 1990.
- Nr. 32: ZELEWSKI,STEPHAN: Schwache Argumente gegen die starke KI-These - Ein Beitrag zur Diskussion über Searle's chinesisches Zimmer, Köln 1990.