Chinesische Zimmer, Turnhallen und Gehirne

Ein wenig Kritik an Searle und zweien seiner Kritikerinnen*

Christopher von Bülow†

4. Oktober 1990

Abstract

I present John Searle's arguments against the 'strong Artificial Intelligence thesis' (according to which a computer that can pass the Turing test can think in the same sense a person can) and counterarguments by Patricia and Paul Churchland. Along the way, I submit my own criticisms of these arguments.

To show that the strong AI thesis is false, Searle gives a thought experiment and a 'formal proof'. In his Chinese-room thought experiment he imagines himself instantiating an AI program that allegedly understands Chinese. Since instantiating the program doesn't make *him* understand Chinese, neither, he infers, can it do so for a computer. The Churchlands object that, while serial-processing computers indeed could not understand Chinese, neural networks could, because of their superior efficiency. I disagree with this putative contrast in ability: serial processors can do anything parallel processors can; that they are slower is inessential. My own criticism of Searle's reasoning is a version of the 'system objection', which says that whereas the processor (Searle) by himself does not understand Chinese, the whole system consisting of the processor, the storage facilities and the program installed does. To rebut the system objection, Searle modifies his thought experiment such that he learns the whole program by rote; thus he *is* the system but he still doesn't understand Chinese. I counter by reference to the hierarchy of levels of abstraction involved.

Searle's formal proof goes from premises involving the notions of syntax and semantics to the conclusion that computer programs are neither constitutive nor sufficient for mind. Here, I agree with the Churchlands' debunking of the proof: they construct a parallel 'proof' that purports to refute Maxwell's theory of electromagnetism. I demonstrate that Searle's proof relies on equivocations on the terms "syntax" and "semantics".

I also deal with Searle's contentions that the difference in 'causal powers' between brains and computers is important for mind, and that programs, having no intrinsic semantics, can be arbitrarily reinterpreted.

*Ich verwende im folgenden stets die weibliche Form, wenn ich von Personen unbestimmten Geschlechts spreche. Die Männer sind natürlich immer mitgemeint.

†eMail: Christopher.von.Buelow@uni.kn; Website: www.uni.kn/FuF/Philo/Philosophie/philosophie/index.php?article_id=88.

I'm sitting here in the abandoned brain,
Waiting for take-off in it.
They say it's never gonna work again,
But I can spare a few minutes.
Been here before in the abandoned brain,
There's flowers on all the controls.
The tape keeps telling me again and again
That I'm the keeper of souls.

The wind blows hard on the abandoned brain,
But there's nobody thinking at all.
The hypothalamus is open to the rain
And the leaves sweep into the hole.
There's no one else in the abandoned brain,
But that's not necessarily bad.
It feeds on itself, but it's not insane;
This brain's too old to go mad.

Roses bloom in the abandoned brain
And thoughts run wild on the floor.
Like a headless corpse, a derailed train;
Who could ask for anything more?
I'm sitting here in the abandoned brain,
Waiting for take-off in it.
They say it's never gonna work again,
But I can spare a few minutes.

Robyn Hitchcock: The Abandoned Brain

Einleitung

Dieser Aufsatz stützt sich auf zwei Artikel, die 1990 in der Märzausgabe der Zeitschrift *Spektrum der Wissenschaft* (bzw. in der Januarausgabe von *Scientific American*) erschienen.[1] Der erste trägt die Überschrift „Ist der menschliche Geist ein Computerprogramm?" und stammt von John R. Searle, der an der Universität von Kalifornien in Berkeley Philosophie lehrt. Searle vertritt darin die These, daß Denken[2] nicht allein durch das Ablaufenlassen von Computerprogrammen entstehen kann. Um diese These zu untermauern, führt er sein bekanntes *Chinese Room*-Gedankenexperiment sowie einen sogenannten ‚formalen Beweis' an.

Der andere Artikel hat den Titel „Ist eine denkende Maschine möglich?" und ist die Entgegnung von Patricia Smith Churchland und Paul M. Churchland auf Searles Artikel. Die Churchlands sind ebenfalls Philosophieprofessorinnen in Kalifornien, und zwar an der Universität San Diego. Sie fechten Searles These an, indem sie versuchen, seinen Beweis ad absurdum zu führen, und den Computern herkömmlicher Bauart Elektronenrechner gegenüberstellen, deren funktionale Architektur an die tierischer Gehirne angelehnt ist, sogenannte *neuronale Netze*. Letztere seien zum einen

[1] Die Seitenangaben beziehen sich i. a. auf die deutschsprachigen Artikel, soweit der Kontext nichts anderes impliziert.

[2] oder *Geist* oder *Bewußtsein* oder *Intelligenz* oder *Intentionalität* oder *mentale Inhalte* oder *Semantik* (in *Searles* Sinn): Ich verwende diese Begriffe als äquivalent in dem Sinne, daß ich davon ausgehe, daß einem ‚System' (Mensch, Tier, Marsmensch, Computer, Programm, Thermostat – irgendeinem *Objekt* also) einer davon genau dann zugeschrieben werden kann, wenn ihm irgendein anderer davon zugeschrieben werden kann.

herkömmlichen, seriell arbeitenden Rechnern auf bestimmten Gebieten weit überlegen und zum anderen ohnehin nicht von Searles ‚Beweis' betroffen, da dieser sich nur auf Rechner beziehe, die nach rein formalen Regeln Symbole manipulieren, und neuronale Netze gar nicht zu diesen zu zählen seien.

Es kann nicht schaden, vor der Lektüre dieses Aufsatzes die beiden Artikel zu lesen. Mein Ziel ist es hier, die Hauptaussagen der Churchlands zu referieren, auf die Mängel ihres Artikels, wie ich sie sehe, hinzuweisen und Searles Behauptungen zu widerlegen oder zumindest zu entkräften.

Der Aufsatz beruht auf einem Referat, das ich im Sommersemester 1990 im Rahmen des Seminars *Probleme des Verhältnisses von Geist und Materie* bei Dr. Martin Carrier, Prof. Dr. Andreas Elepfandt und Prof. Dr. Gereon Wolters an der Universität Konstanz gehalten habe.

Die starke KI-These

Searle unterscheidet eine schwache und eine starke KI-These.[3] Die *schwache KI-These* besagt, Computermodelle seien brauchbare Hilfsmittel beim Studium des menschlichen Geistes, wohingegen die *starke KI-These* besagt, daß ein Programm, das den menschlichen Geist so gut simuliert, daß es den *Turing-Test*[4] besteht, selbst ein denkender Geist im gleichen Sinne wie der menschliche Geist ist.

Diese starke KI-These versucht Searle zu widerlegen, während die Churchlands angetreten sind, sie (oder zumindest eine ähnliche Behauptung) zu verteidigen. Als eine Begründung für ihren Optimismus (oder jedenfalls den Optimismus der KI-Pionierinnen) hinsichtlich der starken KI-These geben die Churchlands Churchs These und ein Resultat Turings an. *Churchs These* besagt, daß jede effektiv berechenbare Funktion rekursiv berechenbar ist,[5] und Turing hat bewiesen, daß jede rekursiv

[3] *KI* steht für *Künstliche Intelligenz*.

[4] Dieser von Alan M. Turing erdachte Test besteht darin, daß eine menschliche Befragerin schriftlich (z. B. über zwei Computerterminals) mit einem KI-Programm und mit einem Menschen kommuniziert. Dabei weiß die Befragerin zwar, *daß* hinter einem Terminal ein Programm steckt, nicht aber, hinter *welchem*. Es können beliebige Themen (z. B. Kindheitserlebnisse, Poesie, Politik) behandelt werden. Das Programm soll nun die Befragerin über seinen künstlichen Ursprung ‚hinwegtäuschen', während die menschliche Kommunikationspartnerin sich bemüht, die Befragerin die wirkliche Situation erkennen zu lassen. ‚Unsichtbar' hinter ihren Terminals verborgen versuchen also Mensch wie Programm einen möglichst menschlichen Eindruck bei der Befragerin zu erwecken. Nach längerem, eingehenden Sondieren muß die Befragerin angeben, hinter welchem Terminal sie den Menschen und hinter welchem sie das Programm vermutet. Diese Befragung wird nun mehrere Male mit wechselnden Befragerinnen und wechselnden menschlichen Kommunikationspartnerinnen durchgeführt (bevorzugt erfahrene KI-Gegnerinnen – aber diese Ergänzung stammt von mir), wobei die Befragerin die menschliche Befragte nicht kennen sollte (ebenfalls meine Ergänzung); hundert wäre in meiner Vorstellung eine größenordnungsmäßig akzeptable Zahl von solchen ‚Simultaninterviews'. Wird das Programm in etwa der Hälfte der Fälle nicht entlarvt, so hat es den Turing-Test bestanden. (Turing verlangt nur, daß das KI-Programm prozentual wenigstens ebensooft für einen Menschen gehalten wird, wie in der analogen Situation ein Mann, der sich als Frau ausgibt, für eine Frau gehalten wird.) Der Turing-Test ist statistischer Natur, so daß es keinen Punkt (keine bestimmte Zahl von Interviews) gibt, ab dem er endgültig bestanden wäre, sondern gewissermaßen nur ‚Grade' des Bestehens. Auch ist das ‚hochgradige' Bestehen des Turing-Tests kein philosophisch zwingender Beweis dafür, daß das Programm tatsächlich Bewußtsein besitzt. Es beweist lediglich, daß das Programm sich *verhalten* kann wie ein Mensch, daß es ebenso flexibel und vielseitig ist wie ein Mensch (s. auch Turing 1950). Bedauerlich ist, daß Searle den Turing-Test sehr verzerrt darstellt, fast als hätte ein Programm ihn schon dann bestanden, wenn es auf Knopfdruck „Ich liebe dich!" ausdrucken kann. Die Churchlands beschreiben den Turing-Test allerdings auch nicht in einer Weise, die ihm gerecht würde.

[5] *Rekursive Berechenbarkeit* ist ein mathematischer Begriff, der besagt, daß eine Funktion in gewisser einfacher Weise aus wenigen, fest vorgegebenen, einfachen Funktionen ‚zusammensetzbar' und in diesem Sinne selbst ‚einfach' ist. Eine Funktion ist *effektiv berechenbar*, wenn ihr Wert für beliebige Argumente mittels irgendeines eindeutig festgelegten Algorithmus in endlicher Zeit berechenbar ist. Der Begriff des

berechenbare Funktion Turing-berechenbar[6] ist. Es ist damit sehr wahrscheinlich, daß jede effektiv berechenbare Funktion von Computern berechnet werden kann. Nun könnte man im Prinzip auch das Verhalten einer Person als Funktion betrachten: Ihre Wahrnehmungen werden als *Argument*, ihre Reaktionen darauf als *Wert* dieser Funktion betrachtet.[7] Die Churchlands schließen daraus, „daß eine geeignete SM-Maschine[8] diese (vermutlich effektiv berechenbare) Funktion berechnen könnte, wie immer sie auch aussehen mag“ (S. 48, Mitte). Dabei erwähnen sie ein wichtiges Problem, ob nämlich das Verhalten eines Menschen prinzipiell in endlicher Zeit durch einen eindeutig festgelegten Algorithmus berechenbar ist, nur ganz am Rande, als sei es keiner Überlegung wert.

Das Chinesische Zimmer

Eines von Searles zwei Hauptargumenten gegen die starke KI-These ist das Gedankenexperiment vom Chinesischen Zimmer, das er wie folgt beschreibt:

> Nehmen Sie eine Sprache, die Sie nicht verstehen. Ich persönlich verstehe kein Chinesisch; für mich sind chinesische Schriftzeichen nur sinnlose Krakel. Stellen Sie sich nun vor, ich würde in ein Zimmer gesetzt, das Körbe voller Kärtchen mit chinesischen Symbolen enthält. Nehmen wir ferner an, man hätte mir ein Buch in meiner Muttersprache Englisch in die Hand gedrückt, das angibt, nach welchen Regeln chinesische Zeichen miteinander kombiniert werden. Dabei werden die Symbole nur anhand ihrer Form identifiziert, ohne daß man irgendeines verstehen muß. Eine Regel könnte also sagen: „Nimm ein Krakel-Krakel-Zeichen aus dem Korb Nummer 1 und lege es neben ein Schnörkel-Schnörkel-Zeichen aus Korb

Algorithmus ist jedoch nicht klar definiert, sondern setzt ein intuitives Verständnis voraus, so daß auch der Begriff der effektiven Berechenbarkeit kein klar umrissener mathematischer Begriff ist. Daher ist Churchs These nicht mathematisch beweisbar. Alle Ergebnisse der Berechenbarkeitstheorie deuten jedoch darauf hin, daß sie zutrifft.

[6] Eine Funktion ist *Turing-berechenbar*, wenn eine Turing-Maschine existiert, die sie berechnet. Dabei sind Turing-Maschinen sehr primitive mathematische Modelle für Computer, die aber im Prinzip alles können, was Computer können. (Allerdings würden physikalische Realisierungen einer Turing-Maschine i.a. unglaublich langsam im Vergleich zu einem normalen Computer sein.) Eine Funktion ist also Turing-berechenbar genau dann, wenn sie durch ein herkömmliches Computerprogramm berechnet werden kann.

[7] Man mag sich fragen, wie denn menschliche Wahrnehmungen für einen Computer ‚verdaulich‘ sein könnten, und wie ein Computer am Ende der Berechnung einer hypothetischen ‚Personenfunktion‘ menschliche Reaktionen ‚ausspucken‘ könnte: Es wäre z.B. denkbar (wenn auch nicht unbedingt machbar), alle von den menschlichen Sinnesorganen zum Gehirn hereinströmenden Nervenimpulse zu messen und als Zahlenfolge zu codieren und dann den Computer so zu programmieren, daß er aus diesen Zahlen andere Zahlen berechnet, die in ähnlicher Weise decodiert, d.h. als vom Gehirn über die Nerven an die Organe ausgehende Befehlsfolge interpretiert werden können. – Die Churchlands schreiben: „Natürlich kennt zum gegenwärtigen Zeitpunkt niemand die Funktion, die das Ausgabeverhalten einer mit Bewußtsein ausgestatteten Person erzeugen würde“ (S. 48, Mitte), als ginge es bei der KI um die Berechnung *einer bestimmten* Funktion, mit der man dann das Wesen des Geistes erfaßt hätte, d.h. als würden alle intelligenten Wesen in gleichen Situationen gleich reagieren. Weiter scheinen sie zu glauben, es genüge, den Zustand der Umwelt einer Person festzulegen, um auch den Wert ihrer ‚Eingabe–Ausgabe-Funktion‘ festzulegen. Es ist aber doch so, daß die Reaktion einer Person auf ihre Umwelt sehr stark abhängig von ihrem inneren Zustand ist, so daß auf gleiche Wahrnehmungen selten gleiche Reaktionen folgen würden. – Genaugenommen wäre eine solche ‚Personenfunktion‘ auch weniger eine mathematische Funktion als eine mathematische *Maschine*, denn es würde von ihr nicht erwartet werden, für ein *einzelnes* Argument (einen Wahrnehmungskomplex) jeweils einen *einzelnen* Wert (eine Reaktion) zu liefern, sondern einen kontinuierlichen Strom von Eingaben zu schlucken und *gleichzeitig* einen kontinuierlichen Strom von Ausgaben zu liefern. Menschen wechseln ja nicht ab zwischen wahrnehmen und reagieren, sondern tun beides dauernd.

[8] Eine Maschine, die Symbole nach formalen Regeln manipuliert; kurz: eine **S**ymbole **m**anipulierende Maschine, d.h. ein Computer.

> Nummer 2.“ Angenommen, von außerhalb des Zimmers würden mir Menschen, die Chinesisch verstehen, kleine Stöße von Kärtchen mit Symbolen hereinreichen, die ich nach den Regeln aus dem Buch manipuliere; als Ergebnis reiche ich dann meinerseits kleine Kartenstöße hinaus. In die Computersprache übersetzt wäre also das Regelbuch das Computerprogramm, sein Autor der Programmierer und ich der Computer; die Körbe voller Symbole wären die Daten, die kleinen mir ausgehändigten Stöße die Fragen und die von mir hinausgereichten Stöße die Antworten. Nehmen wir nun an, das Regelbuch sei so verfaßt, daß meine Antworten auf die Fragen von denen eines gebürtigen Chinesen nicht zu unterscheiden sind. Beispielsweise könnten mir die Leute draußen eine Handvoll Symbole hereinreichen, die – ohne daß ich das weiß – bedeuten: „Welches ist ihre Lieblingsfarbe?“ Nach Durcharbeiten der Regeln würde ich dann einen Stoß Symbole zurückgeben, die – was ich ebensowenig weiß – beispielsweise hießen: „Meine Lieblingsfarbe ist blau, aber grün mag ich auch sehr.“ Also hätte ich den Turing-Test für Chinesisch bestanden. Gleichwohl habe ich nicht die geringste Ahnung von dieser Sprache. Und ich hätte auch keine Chance, in dem beschriebenen System Chinesisch zu lernen, weil es mir keine Möglichkeit bietet, die Bedeutung irgendeines Symbols in Erfahrung zu bringen. Wie ein Computer hantiere ich mit Symbolen, aber verbinde keine Bedeutung mit ihnen. Der Punkt des Gedankenexperiments ist der: Wenn ich kein Chinesisch verstehe, indem ich lediglich ein Computerprogramm zum Verstehen von Chinesisch ausführe, dann tut das auch kein Digitalcomputer. Solche Maschinen hantieren nur mit Symbolen gemäß den im Programm festgelegten Regeln. Was für Chinesisch gilt, gilt für andere geistige Leistungen genauso. Das bloße Hantieren mit Symbolen genügt nicht für Fähigkeiten wie Einsicht, Wahrnehmung, Verständnis oder Denken. Und da Computer ihrem Wesen nach Geräte zur Manipulation von Symbolen sind, erfüllt das bloße Ausführen eines Computerprogramms auch nicht die Voraussetzungen einer geistigen Tätigkeit. (S. 40 f)

Ich werde im Folgenden noch mehrmals auf dieses Gedankenexperiment zurückkommen.

Die konnektionistische Alternative und das Chinesische Zimmer

Trotz ihres Optimismus sehen die Churchlands die Schwierigkeiten der klassischen (d. h. auf herkömmliche, seriell arbeitende Computer gestützten) KI: Klassische Programme bewältigen manche Spezialistinnenaufgaben (z. B. Schach, logische Ableitungen) gleich gut wie oder besser als menschliche Expertinnen, versagen aber bei vielen Problemstellungen, die Menschen dauernd schnell und mühelos bewältigen (z. B. die Verarbeitung optischer und akustischer Signale (Mustererkennung) oder bestimmte Gedächtnisleistungen). Gerade auf letzterem Gebiet glänzen jedoch Computer(-programme) eines neuen Typs, die sogenannten *neuronalen Netzwerke*, in denen Daten nicht nacheinander (*seriell*) von einem einzigen zentralen Prozessor, sondern mehr oder weniger gleichzeitig (*parallel*) verteilt in vielen miteinander vernetzten (deshalb *Konnektionismus*) einfachen Prozessoren verarbeitet werden. Aufgrund dieser andersartigen funktionalen Architektur können neuronale Netzwerke sehr schnell bestimmte sehr komplexe Aufgaben lösen, die herkömmliche Computer, wenn überhaupt, so nur entweder mangelhaft oder sehr langsam bewältigen.

Nach Meinung der Churchlands sind also die Aussichten der klassischen KI schlecht (wenn auch herkömmliche Computer für viele Anwendungsgebiete weiterhin unabdingbar bleiben werden), wohingegen die neue, neuronale KI zu großen Hoffnungen Anlaß gibt.

Sie sind daher gern bereit zuzugeben, daß Searles Chinesisches Zimmer kein Chinesisch versteht. Sie bemängeln, daß das Chinesische Zimmer „mit absurder Langsamkeit arbeitet“ und nennen es ein „Rube Goldberg system“ (S. 28, links) (in der deutschen Übersetzung: „Karl-Valentin-System“ (S. 50, links)). Es leidet also in ihrer Vorstellung an der Krankheit der seriellen Prozessoren: Da sich die funktionale Architektur dieses ‚Computers‘ radikal von der des Gehirns unterscheidet, kann er auf dessen Terrain nicht mit dem Gehirn wetteifern. Wenn man einen so ungeeigneten Prozessor wählt, ist es ihrer Ansicht nach kein Wunder, wenn kein Geist entsteht.

Searle wendet dagegen m. E. zu Recht ein, daß einerseits Geschwindigkeit für das Gedankenexperiment keine Rolle spielt und andererseits neuronale Netze sich, soweit es seine Überlegungen betrifft, nicht von herkömmlichen Computern unterscheiden. Diese Einwände möchte ich unterstützen.

Was das Gedankenexperiment angeht, so stimme ich mit Searle darin überein, daß die Geschwindigkeit, mit der er im Chinesischen Zimmer Antworten produziert, unwesentlich für die Frage ist, ob irgendwo im Zimmer Chinesisch verstanden wird. Auch wenn die Antworten immer erst Jahre nach den Fragen kämen, würde ich dem Programm deswegen nicht das Verständnis oder die Intelligenz absprechen. Natürlich könnte das Programm so nie und nimmer den Turing-Test bestehen. Aber für das Gedankenexperiment besitzen wir die Freiheit, Searle beliebig schnell mit seinen Kärtchen hantieren zu lassen, unabhängig davon, wie zermürbend langsam dieser ‚Computer‘ sicher wäre, würde man das Gedankenexperiment tatsächlich in die Realität umsetzen. Ebenso können wir beliebige Fortschritte in der Computertechnologie postulieren, um hypothetische serielle Prozessoren an Rechengeschwindigkeit mit jedem möglichen neuronalen Netz gleichziehen zu lassen. Die Beweiskraft des Gedankenexperiments wird dadurch nicht geschwächt. Wie *schnell* ein Computer formale Symbole manipuliert, ändert nichts an der Tatsache, *daß* er dies tut.

Da also m. E. der konnektionistische Ausweg gar kein Ausweg ist, bleibe ich im Gegensatz zu den Churchlands dabei, dem Chinesischen Zimmer Geist zuzusprechen. Ich werde dies später begründen und erläutern, weswegen meiner Ansicht nach Searles Gedankenexperiment die starke KI-These nicht widerlegt.

Bezüglich der neuronalen Netzwerke sagt Searle ganz richtig, daß alles, was sie können, auch herkömmliche Computer können. Man muß sie nur dahingehend programmieren, daß sie ein neuronales Netz simulieren (bzw. *emulieren*), indem sie dessen parallele Berechnungen seriell ausführen und die Wirkungsweise der Vernetzung getreu nachahmen. Eine solche Programmierung ist unproblematisch und bei den ‚Neuroinformatikerinnen‘ gang und gäbe.

Nun gibt es (auch aus den Reihen der KI-Befürworterinnen) Stimmen, die meinen, die Simulation eines neuronalen Netzes sei nicht wirklich ein neuronales Netz, sondern eben nur eine Simulation, so daß die berechnungstheoretische Gleichwertigkeit der herkömmlichen Computer mit den Netzen nicht auch eine Gleichwertigkeit in Bezug auf das Mentale nach sich zöge. D. h. wenn ein *echtes* neuronales Netz intelligent ist, so noch lange nicht die Simulation dieses Netzes auf einem seriellen Rechner. Dies ähnelt verdächtig Searles Worten über Duplikation und Simulation von Geist bzw. Gehirnen (s. „Die kausalen Kräfte von Gehirnen und Programmen“ auf Seite 15). Es überrascht als Haltung von KI-Befürworterinnen besonders insofern, als doch gerade die starke KI-These impliziert, daß die Simulation der dem Geist zugrundeliegenden physikalischen Vorgänge eine Duplikation des Geistes selbst bedeuten könne. Und da soll die Simulation der in einem Parallelcomputer stattfindenden Rechenprozesse nicht dessen eventuelle geistige Leistungen duplizieren können?

Allerdings vertreten einige KI-Befürworterinnen nur eine abgeschwächte Version

der starken KI-These, wonach es bei KI-Programmen nicht nur darauf ankommt, *was* sie leisten, sondern auch darauf, *wie* sie es bewerkstelligen. Dieser Standpunkt muß jedoch m. E. nicht die unterschiedliche Bewertung von echten neuronalen Netzen und ihren Simulationen nach sich ziehen. Schließlich sind die Rechenprozesse, die in einer sauberen *Simulation* eines Netzes ablaufen, auf den relevanten Ebenen isomorph zu denen, die bei gleichem Input in diesem Netz *selbst* ablaufen.

Die Churchlands scheinen wirklich nicht an die ‚mentale Gleichwertigkeit' von neuronalen Netzen und ihren Simulationen zu glauben. Sie sagen, daß neuronale Netze *nicht* nach struktursensitiven Regeln Symbole manipulieren. „Regelgesteuerte Symbolmanipulation scheint vielmehr nur eine von vielen kognitiven Fähigkeiten zu sein, die ein Netz erlernen kann oder auch nicht; sie ist jedenfalls nicht sein grundlegender Modus operandi" (S. 52 f). Klar ist aber doch, daß erstens neuronale Netze ihren Input nur von seiner Form abhängig bearbeiten und zweitens die Leistungen neuronaler Netze nur auf der formalen Symbolmanipulation ihrer Knoten beruhen, auch wenn man die Art, in der das gesamte Netz das Eingabe‚symbol' (den Eingabevektor) verarbeitet, als formale Symbolmanipulation zu bezeichen sich sträubt.

Die Chinesische Turnhalle

Searle begegnet dem konnektionistischen Argument zusätzlich, indem er sein Gedankenexperiment abwandelt: Statt des Chinesischen Zimmers soll man sich eine Turnhalle vorstellen, in der viele nur Deutsch sprechende Menschen die Knoten und Verbindungen eines neuronalen Netzes (bzw. die Neuronen und Synapsen eines chinesischen Gehirnes) nachahmen. Während vorher entsprechend dem herkömmlichen Computermodell *ein* Prozessor die gesamte Datenverarbeitung durchführte (nämlich Searle), haben wir nun viele Personen, die jeweils die Arbeit eines der einfachen Prozessoren im Netz leisten. Und immer noch versteht von diesen Personen keine ein Wort Chinesisch.

Während die Churchlands Searle im Chinesischen Zimmer keine höhere Arbeitsgeschwindigkeit zuzugestehen bereit waren, können sie sich die Chinesische Turnhalle problemlos zur richtigen Größe aufgeblasen vorstellen: Wollte man so z. B. ein menschliches Gehirn mit einer Person pro Neuron und einem Kind für jede synaptische Verbindung simulieren, so würde

> dieses System [. . .] die gesamte menschliche Bevölkerung von mehr als 10 000 Erden erfordern. In einer Turnhalle gäbe das einiges Gedränge. Würde ein solches System dagegen in den geeigneten kosmischen Maßstäben verwirklicht – mit allen Verbindungen wie im menschlichen Fall –, erhielten wir vielleicht ein monströses, langsames, merkwürdig konstruiertes, aber dennoch funktionsfähiges Gehirn. Was seine geistige Leistungsfähigkeit angeht, so läge es sicherlich näher zu vermuten, daß das Monstrum – mit passenden Eingaben versorgt – tatsächlich denken könnte, als daß es nicht dazu imstande wäre. (S. 53, Mitte und rechts)

Hier scheint also die Geschwindigkeit plötzlich keine Rolle mehr zu spielen. Entscheidend ist für die Churchlands also nur, daß der Prozessor die richtige funktionale Architektur hat, so abstrus er ansonsten sein mag. Im Lichte meiner vorhergehenden Überlegungen zur Gleichwertigkeit serieller und paralleler Datenverarbeitung (s. „Die konnektionistische Alternative und das Chinesische Zimmer" auf Seite 5) kann diese Haltung nicht überzeugen.

Searles ‚formaler Beweis'

Wenn sie auch akzeptieren, daß das Chinesische Zimmer kein Chinesisch versteht, kritisieren die Churchlands doch immerhin Searles zweites Hauptargument gegen die starke KI-These, seinen ‚formalen Beweis', der die folgende Gestalt hat:

Axiom 1: Computerprogramme sind formal (syntaktisch).

Axiom 2: Dem menschlichen Denken liegen geistige Inhalte (Semantik) zugrunde.

Axiom 3: Syntax an sich ist weder konstitutiv noch hinreichend für Semantik.

Folgerung 1: Programme sind weder konstitutiv noch hinreichend für Geist.

Das dritte Axiom, mit dem der ‚Beweis' steht und fällt, sei zwar plausibel, aber durchaus nicht zwingend. Searle verlasse sich hier auf den gesunden Menschenverstand. Später werde ich erläutern, weshalb in meinen Augen Searles Beweis keiner ist (s. „Searles drittes Axiom" auf Seite 9).

Zur Unterstützung ihrer Argumentation geben die Churchlands einen analogen ‚Beweis' an, wie ihn im 19. Jahrhundert eine Skeptikerin hätte führen können, die Maxwells These der elektromagnetischen Natur des Lichtes widerlegen wollte.

Axiom 1: Elektrizität und Magnetismus sind Kräfte.

Axiom 2: Die wesentliche Eigenschaft von Licht ist Helligkeit.

Axiom 3: Kräfte an sich sind weder konstitutiv noch hinreichend für Helligkeit.

Folgerung: Elektrizität und Magnetismus sind weder konstitutiv noch hinreichend für Licht.

Wir haben dabei folgende Entsprechungen:

Elektrizität und Magnetismus	Programme
Kräfte	Syntax
Licht	Denken
Helligkeit	Semantik

Die Schlußweise der Skeptikerin ist dabei genau parallel zu der Searles, ihre Axiome sind nach damaligem Wissensstand genauso plausibel wie die Searles nach heutigem; ihre Folgerung hat sich inzwischen jedoch als falsch erwiesen. Es resultiert, daß Searles Folgerung im Laufe der Zeit gleichfalls widerlegt werden könnte, so plausibel sie heute sein mag. Hinsichtlich dieses Pseudobeweises redet sich Searle darauf hinaus, die Schlußweise sei keineswegs analog zu seiner eigenen: Im Falle des Pseudobeweises gehe es um physikalische Kausalität, wohingegen man es beim Chinesischen Zimmer nicht mit Kausalität zu tun habe, da „formale Symbole [. . .] keine kausalen Kräfte im physikalischen Sinne besitzen" (S. 45, Mitte). Dieses Argument finde ich ziemlich fadenscheinig. Ich werde später noch auf Searles ‚kausale Kräfte' eingehen (s. „Die kausalen Kräfte von Gehirnen und Programmen" auf Seite 15).

Das Erleuchtete Zimmer

Die Churchlands lassen weiter ihre fiktive Skeptikerin wie Searle ein Gedankenexperiment anstellen, mit dem sie illustrieren, wie sie die Bedeutung des Chinesischen Zimmers einschätzen:

> „Stellen Sie sich einen dunklen Raum vor, in dem sich ein Mann befindet, der einen Stabmagneten oder einen elektrisch geladenen Gegenstand in der Hand hält. Wenn der Mann den Magneten auf- und abbewegt, dann müßte dieser nach Maxwells Theorie der Künstlichen Helligkeit (KH) einen sich ausbreitenden Kreis elektromagnetischer Wellen und damit Helligkeit erzeugen. Aber wie jeder von uns, der mit Magneten oder geladenen Kugeln herumgespielt hat, nur zu gut weiß, produzieren ihre Kräfte (oder irgendwelche anderen Kräfte), selbst wenn sie bewegt werden, keinerlei Helligkeit. Es ist somit unvorstellbar, daß man wirkliche Helligkeit einfach dadurch erzeugen kann, daß man Kräfte umherbewegt!“ (S. 50 f)

Man könnte bezüglich dieses Gedankenexperiments entgegen dem Augenschein darauf beharren, daß in diesem Zimmer durchaus Licht sei, allerdings von zu großer Wellenlänge und zu geringer Intensität, um von Menschen wahrgenommen zu werden. Oder man entgegnet der Skeptikerin, es sei kein Wunder, daß kein elektrisches Licht entstehe, wenn jemand es mit so ungeeigneten Mitteln zu erzeugen versuche. Entsprechend dürfen wir uns die Haltung der Churchlands zum Chinesischen Zimmer vorstellen: auf Mängel an Geschwindigkeit, Intensität, Qualität verweisend.

Searles drittes Axiom

Searle bestreitet, sein drittes Axiom sei nur plausibel; er hält es vielmehr für eine „selbstevidente logische Grundwahrheit“ (S. 45, rechts). Das scheint mir doch etwas übertrieben. Zugegebenermaßen würde auch ich sein drittes Axiom *so, wie es dasteht*, akzeptieren, aber ich denke, daß es so, wie es dasteht, Searles Folgerung nicht rechtfertigt. Genauer gesagt: Searle verwendet in seinem ‚formalen Beweis‘ die Worte *Syntax* und *Semantik* stillschweigend an verschiedenen Stellen auf verschiedene Weise. Üblicherweise beziehen sich diese Worte auf eine Sprache. Beschäftigen wir uns beispielsweise mit der chinesischen Sprache, so können wir, wenn wir die *Syntax* beherrschen, korrekte (aber nicht notwendig sinnvolle) Sätze bilden. Beherrschen wir die *Semantik*, so verstehen wir chinesische Worte und sinnvolle chinesische Sätze, d. h. wir kennen ihre *Bedeutung*.[9]

Faßt man *Syntax* und *Semantik* so auf (und das dritte Axiom suggeriert diese Lesart), dann ist Axiom 3 durchaus akzeptabel – aber der Beweis funktioniert nicht. Erstens bedeutet *Semantik* in Axiom 2 durchaus nicht Semantik im obenbeschriebenen Sinne. Denkt man an das zum ‚Beweis‘ gehörige Gedankenexperiment, so geht es in Axiom 2 um die *Semantik des Chinesischen* in einem ganz anderen Sinne, nämlich um ein mentales Phänomen, um „geistige Inhalte“.[10] Zweitens: Beziehen wir Axiom 2 beispielsweise auf die Semantik (im üblichen oder in Searles mentalem Sinn) der chinesischen Sprache, so geht das noch lange nicht auch in Axiom 1 mit der Syntax. Angenommen nämlich, das von Searle im Chinesischen Zimmer ausgeführte Programm beinhaltete nichts als die Syntax des Chinesischen. Dann wäre es zwar fähig, grammatikalisch korrekte Sätze zu bilden, kaum aber, auf chinesische Fragen so sinnvolle Antworten zu geben, daß es den Turing-Test bestünde. Dieses Programm muß also erheblich mehr beinhalten. Eine Möglichkeit wäre, daß das Programm die Vorgänge im Gehirn einer Chinesin während einer auf Chinesisch geführten Unterhaltung auf der Neuronenebene simuliert. Dies könnte man, will man bei dem Wort

[9] *Semantik*, bezogen auf eine Sprache, hat noch nicht zwangsläufig etwas mit Geist zu tun. Die Semantik einer Sprache könnte z. B. auch eine mathematische Funktion sein, die einem Wort den Gegenstand und einem Satz den Sachverhalt, für den er steht (oder vielleicht auch seinen Wahrheitswert), zuordnet.

[10] Dies macht den Beweis zwar unsauber, schwächt ihn jedoch noch nicht unbedingt, denn man wird gerne zugestehen, daß ein Programm, wenn es noch nicht einmal die herkömmliche Semantik des Chinesischen hervorbringen kann, erst recht nicht dessen Semantik in Searles Sinn hervorbringen kann.

Syntax bleiben, als *Syntax eines chinesischen Gehirns*[11] bezeichnen. Searle gebraucht also erstens in den Axiomen 1 und 2 die Worte *Syntax* und *Semantik* nicht im von Axiom 3 suggerierten herkömmlichen Sinn, und zweitens haben die Worte *Syntax* und *Semantik* in Axiom 1 und 2 verschiedene Bezugsgegenstände, d.h. wir haben es nicht mit Syntax und Semantik *derselben Sache* zu tun.

Wenn wir Searle nun aber zugestehen, diese Worte in einem etwas weiteren Sinne zu benutzen, dann bleibt sein drittes Axiom auf der Strecke. Zwar ist es sehr plausibel, daß die Syntax des Chinesischen dessen Semantik *nicht* hervorbringen kann, aber ob dazu beispielsweise die ‚Syntax' eines ganzen *chinesischen Gehirns* in der Lage ist, das ist eine ganz andere Frage. Für mich ist die Behauptung, ein den neuronalen Prozessen im Gehirn exakt nachmodellierter Haufen Symbolmanipuliererei in einem Computer könne ein Verständnis des Chinesischen bewirken, genauso plausibel wie die Behauptung, ein verzwicktes Netz von neurophysiologischen Abläufen in den Zellen eines Gehirns könne dies. Denn ich kann mir den Quantensprung von physikalischen Prozessen zum Geist im Falle des Gehirns genausowenig vorstellen wie im Falle des Computers. Und doch ist ersterer offensichtlich möglich. Searles ‚Beweis' verdankt also seinen Anschein von Stringenz dem unsauberen Gebrauch der Worte *Syntax* und *Semantik*.

Das Chinesische Zimmer und der System-Einwand

Die Churchlands geben die starke KI-These auf, wenn sie darauf verzichten, die Fähigkeit serieller Rechner zu verteidigen, aufgrund ihrer Programmierung das materielle Fundament für Intelligenz zu sein, und außerdem behaupten, neuronale Netzwerke betrieben nicht unbedingt formale Symbolmanipulation. Die verbleibenden KI-Positionen verteidigen sie jedoch meiner Ansicht nach ohne Erfolg, da sie das Gedankenexperiment mit dem Chinesischen Zimmer weder widerlegt noch wirklich entkräftet haben. Auch ich sehe den konnektionistischen Weg als sehr vielversprechend an, möchte aber darüberhinaus weiterhin die starke KI-These vertreten. Ich werde mich dabei auf eine abgewandelte Version des von Searle behandelten (S. 45, links) System-Einwandes stützen.

Searles Argumentation bezüglich seines Gedankenexperiments verläuft etwa so: Wenn ‚irgendwo' im Chinesischen Zimmer ein Verständnis des Chinesischen wäre, so müßte es bei Searle selbst sein; da er aber *kein* Chinesisch versteht, könne im Chinesischen Zimmer kein Verständnis der chinesischen Sprache und damit auch keine durch das Ablaufenlassen des Programmes erzeugten mentalen Phänomene sein. Das Ablaufenlassen des Programmes sei also nicht hinreichend für das Vorhandensein von Geist.

Um Searles Gedankenexperiment zu entkräften, ist es nicht nötig zu *beweisen*, daß das Programm im Chinesischen Zimmer Chinesisch versteht.[12] Es genügt zu zeigen,

[11] Die Beschreibung dieser neuronalen Vorgänge ist *syntaktisch* insofern, als sie nicht auf die begleitenden mentalen Vorgänge (‚Semantik': die Gedanken und Empfindungen der Chinesin) zurückgreift, sondern lediglich vom physikalischen Geschehen im Gehirn spricht, wo Nervenimpulse nur entsprechend ihren physikalischen Eigenschaften (ihrer ‚Form') den Naturgesetzen (‚syntaktischen Regeln') folgend weitere Nervenimpulse auslösen (‚verarbeitet werden').

[12] M.E. ist es prinzipiell nicht möglich, das Vorhandensein von Geist streng zu beweisen (ausgenommen jeder einzelnen das Vorhandensein ihres eigenen Geistes). Natürlich gehen wir immer davon aus, daß die Menschen, mit denen wir umgehen, Bewußtsein wie wir selbst besitzen, aber theoretisch besteht immer die Möglichkeit, daß sie nur geschickt konstruierte Automaten sind: Wir können in anderer Leute Geist nicht ‚hineingucken'. Es ist also auch nicht möglich, die starke KI-These zu *beweisen*, selbst wenn eines Tages Programme alltäglich sein sollten, die den Turing-Test bestehen. Ob man solchen Programmen Bewußtsein

daß es Chinesisch verstehen könnte, ohne daß Searle davon etwas *bemerkt*. Darauf zielt der *System-Einwand*, der besagt, daß das *Gesamtsystem des Chinesischen Zimmers* (also Searle *und* die Regelbücher *und* die Kärtchen *zusammengenommen*) durchaus Chinesisch verstehen könne, ohne daß dafür *Searle* Chinesisch zu verstehen brauche. In der Tat fände ich es in höchstem Maße wunderbar, wenn Searle im Chinesischen Zimmer irgendetwas vom Inhalt der chinesischen Konversation mitbekäme. Das wäre ja, als ob eine Person, die auf der Hardwareebene betrachtet, was in einem Computer während der Ausführung eines komplizierten Programmes abläuft, verstünde, was der Sinn des Programmes ist; oder als würde eine Person, die auf der neuronalen Ebene betrachtet, was in einem Gehirn vorgeht, daraus ersehen, was dessen Besitzerin denkt. Und Searle erwartet offenbar darüberhinaus, daß in der Situation des Chinesischen Zimmers das Chinesisch-Verständnis des Programmes nicht nur für ihn erfahrbar, sondern zu seinem *eigenen* Verständnis wird.

Searle sieht die Bedeutsamkeit der Unterscheidung zwischen dem Gesamtsystem des Chinesischen Zimmers und dem ‚Subsystem' Searle nicht ein. Er versucht diese Unterscheidung unmöglich zu machen, indem er sein Gedankenexperiment abwandelt: „Stellen Sie sich vor, ich hätte den Inhalt der Körbe [voll Kärtchen mit chinesischen Symbolen] und das Regelwerk auswendig gelernt und würde alle Berechnungen in meinem Kopf durchführen; Sie dürfen sich sogar vorstellen, daß ich unter freiem Himmel arbeite. Es gibt dann nichts an diesem ‚System', das sich nicht in mir befände – und da ich kein Chinesisch verstehe, versteht es auch das System nicht" (S. 45, links).

Ich meine, auch in dieser Situation besteht immer noch ein Unterschied zwischen dem Programm aus dem Chinesischen Zimmer und Searle selbst, auch wenn sich nun beide ‚in Searles Kopf befinden'. Darauf deutet z. B. die Tatsache hin, daß die Information aus dem Programm auf andere Weise Teil von Searle (bzw. seinem Gedächtnis) ist als z. B. seine Englischkenntnisse. Dabei meine ich nicht so etwas wie den Unterschied zwischen stupid auswendiggelerntem Wissen und von frühester Kindheit an erworbener Sprachfähigkeit. Ich ziele vielmehr darauf ab, daß das englische Analogon zum Chinesisch-Programm für Searle unzugänglich im neuronalen Netz seines Gehirns gespeichert ist. Von den Regeln, die zu seinen Englischkenntnissen gehören, kann sich Searle nur einen geringen Teil ins Bewußtsein bringen, denn er kann nicht die Konfiguration der in seinem Gehirn dafür zuständigen Neuronen wahrnehmen. Auch daß er sich dieses geringen Teils bewußt ist, verdankt er nur seiner Bildung, die aber für seine Fähigkeit, Englisch zu verstehen, kaum von Belang ist. (Mit der letzteren Bemerkung meine ich, daß beispielsweise ein Kind durchaus recht gut Englisch verstehen kann, ohne über die Regularitäten, denen sein Englisch-Verständnis unterliegt, irgendetwas von Belang sagen zu können.) Das Chinesisch-Programm hingegen ist ihm in allen Details (aber auch *nur* in den Details, wie ich im nächsten Abschnitt erläutern werde) vollkommen bewußt.

Ich will damit weder irgendwie andeuten, daß man eine Sprache nur dann versteht, wenn man sich ihrer Grammatik bzw. Syntax nicht voll bewußt ist, noch, daß die Regeln, die das Chinesisch-Programm umfaßt, den Regeln der englischen Sprache ähnlich wären, die Searle kennt (etwa daß *Subjekt – Prädikat – Objekt* die Standardreihenfolge in englischen Sätzen ist). (Daß ich Letzteres für unrealistisch halte, habe ich ja schon im vorigen Abschnitt zum Ausdruck gebracht.) Dieser unterschiedliche Grad an ‚Bewußtheit' ist vielmehr nur eine Folge des eigentlich entscheidenden

zuschreibt, bleibt meiner Ansicht nach Glaubenssache. Ich will Searle nicht das Recht abstreiten zu glauben, sie besäßen keines, möchte aber klarstellen, daß dies eben nur *Glauben* und nicht *Wissen* ist, oder jedenfalls, daß keines von Searles Argumenten für diese Behauptung stichhaltig ist.

Unterschiedes.

Den eigentlichen Unterschied zu benennen fehlen mir allerdings buchstäblich die Worte, weswegen ich auf andere Weise versuchen will, ihn zu verdeutlichen. Dabei wird die Idee einer Hierarchie von Abstraktionsebenen[13] nicht nur hilfreich, sondern wesentlich sein. Diese Intuition will ich im Folgenden erläutern.

Hierarchien von Abstraktionsebenen

In der Mathematik gehört es zum täglichen Brot, aus gegebenen Begriffen neue zusammenzusetzen. So kann man etwa mit den Begriffen *Klasse*, *Menge* und *Element* beginnen, um daraus nach und nach beispielsweise die Begriffe des *Paares*, der *Funktion*, der *Mächtigkeit* etc. zu gewinnen. Ähnlich findet in der Informatik eine ‚Evolution' der Programmiersprachen statt: Aufbauend auf schon bekannten (‚implementierten') Sprachen werden neue entwickelt und gebrauchsfertig gemacht. Ganz am Anfang dieser Entwicklung steht dabei die Maschinensprache, die noch sehr nah an die Hardware des Computers angelehnt ist. Das heißt, wenn man in Maschinensprache programmiert, muß man dem Rechner quasi jeden einzelnen noch so simplen ‚Handgriff', den er ausführen soll, genau spezifizieren, wobei die Addition zweier kleiner ganzer Zahlen womöglich schon zu den komplizierteren Akten gehört. Das ist natürlich sehr mühselig, so daß es naheliegt, oft gebrauchte Anweisungsfolgen so zu speichern, daß man sie mit einem einzigen einfachen Befehl als Paket aufrufen kann, ähnlich wie man in der Mathematik, anstatt jedesmal mit einer langen Formel auszudrücken, was man meint, sich dafür ein Wort wie *Funktion* ausdenkt. Dies verringert nicht nur durch Abkürzung den Aufwand, den man treiben muß, um ein bestimmtes Ziel zu erreichen, sondern hilft auch durch Abstraktion, besser zu verstehen, was man eigentlich tut bzw. wovon man spricht (wenn die Befehlspakete – ‚Prozeduren' oder ‚Subroutinen' – bzw. die neuen Begriffe geschickt konstruiert sind). Hat man auf diese Weise eine ausreichende Menge von komplexen ‚Bausteinen' beisammen, so hat man damit eine neue Programmiersprache (bzw. ein neues Feld der Mathematik), und die Benutzerinnen können sich dieser Bausteine bedienen, ohne sich mit dem Wissen belasten zu müssen, wie sie zusammengesetzt sind. So folgt meines Wissens auf die Maschinensprache die Assemblersprache, usw., bis man einige ‚Sprachstufen' höher bei heute gängigen prozeduralen Programmiersprachen wie Pascal, ALGOL 68 oder FORTRAN und schließlich bei (mehr oder weniger, jdf. sogenannten:) deklarativen Sprachen wie z. B. PROLOG angelangt.[14]

Was geht nun im Computer vor, wenn ich beispielsweise ein PROLOG-Programm ablaufen lasse? Das ist eine Frage der Perspektive. Einerseits kann man sagen, PROLOG untersucht, ob aus den Fakten und Regeln in meinem Programm die Aussage folgt, die ich ihm zur Verarbeitung gegeben habe. (So ist PROLOG gedacht: Es geht um Fakten, Regeln und ihre logischen Konsequenzen. Hinter anderen Programmiersprachen stecken ganz andere Ideen.) Dies ist die oberste *Abstraktionsebene*, wo man gewissermaßen die Abläufe im Computer aus der Vogel- bzw. Programmiererinnenperspektive betrachtet. Aber damit der Computer das Programm ausführen kann, muß es (mitsamt den Algorithmen, nach denen der Computer eine Ableitung sucht) letztendlich in die Maschinensprache übersetzt werden, denn der Hardware ist PROLOG nur über diese Vermittlung zugänglich. Man kann also auch sagen, daß der

[13] Diese Idee verdanke ich hauptsächlich den Büchern von Douglas R. Hofstadter, wo man (neben anderen interessanten Gedanken und Bildern) sie wohl besser als hier erklärt findet.

[14] Ich bin mit den ‚niedrigeren' Programmiersprachen nicht vertraut, denke aber, daß meine Darstellung grob korrekt ist.

Computer das Maschinencode-Programm, in das das PROLOG-Programm übersetzt wurde, abarbeitet. Dies wäre die niedrigste informatische Abstraktionsebene, die der Frosch- oder Maschinenperspektive sozusagen.

Zwischen diesen beiden Perspektiven oder Abstraktionsebenen liegt jedoch nicht nur eine ‚Stufe', ein Abstraktions- oder Übersetzungsschritt, sondern viele. Irgendwie muß das Programm die ‚Leiter' von aufeinander aufbauenden Programmiersprachen wieder hinuntertransportiert bzw. -übersetzt werden. Diese Arbeit besorgen *Compiler* oder *Interpreter*, eigene Übersetzungsprogramme. Ich weiß nicht, wie es im Falle konkreter PROLOG-Implementationen aussieht; vorstellbar wäre jedoch, daß das PROLOG-Programm erst einmal in ein Pascal-Programm übersetzt wird, dieses dann in ein Programm in einer wiederum einfacheren Sprache, bis schließlich erst nach mehreren Reduktionsschritten der fertige Maschinencode vorliegt. Jede Zwischenübersetzung bedeutet eine weitere Abstraktionsebene oder Perspektive, aus der die Vorgänge im Computer betrachtet werden können: Jeweils abstrakter und verständlicher als Maschinencode, aber detaillierter und verwirrender als das ursprüngliche PROLOG-Programm. Diese vielen Zwischenstufen existieren nicht von ungefähr, denn jede Stufe steht für einen bewältigbaren Abstraktionsaufwand: Man kann ohne übermäßige geistige Verrenkungen von einer Perspektive zur nächsten wechseln, d.h. etwa in meinem Beispiel Teile des PROLOG-Programms als Pascal-Prozeduren auffassen; aber um den Sprung von einer Abstraktionsebene auf eine mehrere Stufen höher oder tiefer liegende ohne die Vermittlung der dazwischenliegenden Ebenen zu bewältigen, muß man wohl genial sein, wenn es überhaupt möglich ist. Sollten nun eines Tages Programme geschrieben werden, die den Turing-Test bestehen, so wahrscheinlich in Programmiersprachen, die noch einmal etliche Stufen oberhalb der heutigen angesiedelt sein werden.

Mit diesem Bild einer langen Hierarchie verschiedener Perspektiven, aus denen man die Vorgänge in einem Computer betrachten kann, will ich die Idee plausibel machen, daß man denselben Prozeß grundverschieden verstehen und beurteilen kann, abhängig davon, aus welcher Perspektive man ihn betrachtet.

Die Gegenstände und Prozesse, von denen man in der Sprache der obersten Abstraktionsebene spricht, können ganz anderer Natur sein als jene, von denen man in der Sprache der untersten spricht. So können wir im Falle eines wirklich denkenden KI-Programmes von zwei Sorten von Symbolen sprechen; von den Einsen und Nullen, mit denen der Computer (froschperspektivisch betrachtet) hantiert: formalen, nicht bedeutungsbehafteten (jedenfalls nicht im Sinne Searles) Symbolen; und von Vorstellungen und (z.B. chinesischen) Worten, in denen (in der Sprache der Vogelperspektive gesprochen) das Programm denkt:[15] ‚bedeutsamen' Symbolen.[16]

[15] Ich spreche immer davon, daß *Programme* denken, während *Computer* nur formale Symbolmanipulation betreiben. Dem Computer kann man das Programm und damit die Denkfähigkeit wegnehmen, und es bleibt doch derselbe Computer, während das Denken immer da ist, wo das Programm ist, vorausgesetzt, es ist auf geeignete Weise realisiert, d.h. in einem funktionierenden Computer.

[16] In dieser Hinsicht ist die Wahl ausgerechnet der chinesischen Sprache ein geschickter Schachzug: Die formalen Symbole, mit denen Searle hantiert, sind gleichzeitig (wenigstens für die außerhalb des Zimmers befindlichen Chinesinnen) echte, ‚bedeutsame' Symbole. Damit wird suggeriert, die formalen Symbole, mit denen ein Computer hantiert, seien schon (vom gleichen Typ wie) die Symbole, die für den vom Programm hervorgebrachten Geist Bedeutung haben: Worte, Bilder, Vorstellungen, innere Repräsentationen. D.h. Searle erweckt den Anschein, als müßten es die Einsen und Nullen sein, mit denen der Computer hantiert, die für ein KI-Programm eine Bedeutung haben. Suchte man nur an dieser Stelle nach Searles Semantik, so wäre man, selbst wenn sie hier in Erscheinung treten sollte, kaum geneigt, die Anwesenheit eines vollwertigen Geistes/Bewußtseins zu akzeptieren. Man würde jedoch am falschen Ort suchen. Zwar kann man sich die in einer (beispielsweise auf Deutsch geführten) Turing-Test-Konversation auf den Terminal-Bildschirm getippten Worte im Prinzip als in eine Folge von Buchstaben und Satzzeichen oder

Es genügt allenfalls prinzipiell, den Übersetzungsalgorithmus zu kennen und sämtliche Vorgänge im Computer, während ein KI-Programm abläuft, aus der Froschperspektive beobachten zu können. Die tatsächliche Übersetzung dürfte von einem Menschen kaum nachzuvollziehen sein. Praktisch unmöglich ist es, *ohne* Kenntnis des Übersetzungsalgorithmus oder der Wirkungsweise des Programmes (und so armselig steht Searle im Chinesischen Zimmer da) aus der Froschperspektive zu einem Verständnis dessen zu gelangen, was ‚wirklich' geschieht, auf der obersten Abstraktionsebene.

Glaubt man nun an die Möglichkeit der vollständigen Reduktion mentaler Phänomene auf neurophysiologische Vorgänge (und ich vermute, darin stimmen Searle und ich überein), so wird man bereit sein, das soeben für den Computer entwickelte Modell auch für das menschliche Gehirn gelten zu lassen, d.h. die Vorgänge im Gehirn aus einer ähnlichen Vielfalt von Perspektiven zu betrachten wie vorher die im Computer: Auf der untersten Ebene (die ich analog zum Fall des Computers als *Hardwareebene* bezeichnen möchte) feuern Horden von Neuronen und sondern Drüsen schleimige Säfte ab, während auf der obersten Ebene (die ich asymmetrischerweise nicht als *Software-*, sondern als *mentale Ebene*[17] bezeichnen würde) im strahlenden Licht des Geistes die Gedanken und Emotionen umherflitzen. Hier dürfte es noch schwieriger als im Falle von KI-Programmen sein, durch Betrachtung der ablaufenden Prozesse aus der am *einen* Ende der Skala befindlichen Perspektive zu verstehen, was, betrachtet aus der am *anderen* Ende befindlichen Perspektive, vorgeht.[18] D.h. selbst wenn man im Prinzip weiß, wie von der neuronalen oder Hardwareebene aus zu abstrahieren ist, um schließlich deren Phänomene als mentale Phänomene zu verstehen, wird man nicht dazu im Stande sein, diesen Perspektivenwechsel in der einen oder anderen Richtung tatsächlich im Kopf durchzuführen.[19]

Searles Gehirn, während er im Kopf die Berechnungen des auswendiggelernten Chinesisch-Programmes durchführt, ist nun ein besonderer Fall. Hier ist auf die Hierarchie von Abstraktionsebenen von der neuronalen bis zur mentalen Ebene Searles zusätzlich eine weitere ähnliche Skala von Perspektiven aufgepfropft: Deren unterste ist die Hardwareebene eines chinesische Symbole manipulierenden Computers (Searle im Chinesischen Zimmer bzw. Searles chinesisches Oberstübchen), und deren oberste die mentale Ebene des Chinesisch-Programmes. Der Witz ist nun, daß die unterste Ebene der ‚chinesischen Hierarchie' Teil der (bzw. eingebettet in die) oberste(n) Ebene der ‚Searle-Hierarchie' ist. Searle simuliert in Gedanken die Hardware eines Computers, so wie ein Computer einen anderen emulieren kann. Die Einsen und Nullen bzw. die chinesischen Symbole, mit denen der ‚Chinesisch-Computer' hantiert, sind mentale Phänomene, Objekte von Searles Denken.

---

darüberhinaus in eine Folge von Einsen und Nullen zerlegt vorstellen, aber niemand kann erwarten oder verlangen, diese Sprachbausteine hätten für den vom Programm hervorgebrachten Geist irgendeine eigene Bedeutung.

[17] Die Software ist es ja gerade, die aus den erwähnten mannigfaltigen Perspektiven betrachtet werden kann, deren unterste, Hardware-nächste ich *Hardwareebene* nenne. Zwar kann man auch die Hardware beim Computer mit verschiedenen Graden von Abstraktion betrachten, aber dies würde auf die Betrachtung einer Maschine und ihrer Bauteile hinauslaufen, und nicht auf die Betrachtung von *Prozessen*, die für Geist konstitutiv sein können.

[18] Dies deswegen, weil man erwarten darf, daß die Hierarchie von Abstraktionsebenen zwischen Hardwareebene und Geist im Falle des menschlichen Geistes weniger systematisch strukturiert ist als im Falle eines KI-Programmes, da jener ein Produkt der Evolution ist, während dieses das Ergebnis zielgerichteter menschlicher Arbeit wäre.

[19] M.E. ist zwar die Reduktion des menschlichen Geistes auf neurochemische Vorgänge im Prinzip möglich und wird eines Tages durchgeführt werden, aber kaum im Kopf einer einzelnen, sondern im Computer.

Nun endlich können wir wieder die Frage stellen: Warum sollte das Chinesisch-Verständnis des *Programmes* ein Chinesisch-Verständnis bei dem es ausführenden *Searle* bewirken? D.h. warum sollte die Abstraktionsebene, auf der von *Searles* Gedanken zu sprechen sinnvoll ist (die *unterste* der chinesischen Hierarchie), identisch sein mit der, auf welcher die Gedanken des *Programmes* zu den beobachtbaren Objekten gehören (die *oberste* der chinesischen Hierarchie)? Nur weil es auf beiden Ebenen um Verständnis und Geist geht? Die Antwort erübrigt sich. Die Gedanken des Chinesisch-Programmes sind Searle genausowenig durch Introspektion zugänglich (weil in Abstraktionsebenen gemessen ähnlich weit von ‚ihm selbst' entfernt) wie das Feuern seiner Neuronen.

Dies ist *mein* Verständnis des Gedankenexperiments vom Chinesischen Zimmer, und ich halte es für mindestens ebenso akzeptabel wie Searles Verständnis. Meine Vorstellungen über Geist und KI mögen reichlich spekulativ und unausgegoren sein, aber ich behaupte auch nicht, daß sie zutreffen. Es genügt mir, daß meine Intuition bezüglich des Chinesischen Zimmers mindestens so plausibel ist wie die Searles und somit seine nicht die einzig mögliche ist.[20] Damit ist Searles Gedankenexperiment kein zwingendes Argument gegen die starke KI-These mehr.

Die kausalen Kräfte von Gehirnen und Programmen

Die entscheidende Eigenschaft, die in Searles Augen biologische Gehirne KI-Programmen (bzw. Elektronengehirnen) voraushaben, besteht in ihren *kausalen Kräften*. Welche Kräfte das genau sind, die aus Proteinen und dergleichen bestehende Neuronen und die verschiedenen Hormone besitzen und Einsen und Nullen, Additionsanweisungen und Schleifen (bzw. Siliziumchips, Magnetbänder und elektrische Ladungen) nicht, erläutert er nicht genauer, so daß die kausalen Kräfte des Gehirns manchmal ähnlich mystisch wie der göttliche Atemhauch erscheinen. Ich glaube, Searles Haltung so zusammenfassen und beschreiben zu können, daß Geist ein Naturphänomen wie andere auch ist, zu dessen Hervorbringung es nicht genügt, die *Struktur* der es konstituierenden Prozesse nachzuschaffen, sondern darüberhinaus ebendiese Prozesse *selbst* notwendig sind. Statt „kausale Kräfte" sollte man vielleicht einfach immer „physikalische Eigenschaften" lesen. *Nicht* gemeint ist damit jedenfalls die Fähigkeit von Gehirnen (bzw. Computern), über Wahrnehmungsorgane (bzw. ‚Transducer') und ihrer Kontrolle unterworfene Gliedmaßen (ggf. künstliche, so daß der Computer zu einem Roboter wird) mit ihrer Umwelt kausal wechselwirken zu können. Diese Fähigkeit mag für Geist ebenfalls notwendig sein, aber um sie geht es Searle nicht (vgl. „The Robot Reply" in Searle 1980, S. 420). Er meint,

> daß das Gehirn nicht einfach ein formales Muster oder Programm aktiviert (es macht das auch), sondern kraft spezifischer neurobiologischer Prozesse zugleich mentale Ereignisse verursacht. Gehirne sind spezifische biologische Organe, und ihre besonderen biochemischen Eigenschaften befähigen sie, Bewußtsein und andere Formen mentaler Phänomene hervorzurufen. Computersimulationen von Gehirnprozessen liefern dagegen nur Modelle der formalen Aspekte dieser Prozesse. Man sollte Simulation nicht mit Nachschaffen oder Duplikation verwechseln. Das Computermodell mentaler Prozesse ist um nichts realer als ein Computermodell irgendeines anderen natürlichen Phänomens. Man kann sich zum Beispiel eine Computersimulation der Wirkungsweise von Peptiden im Hypothalamus vorstellen, welche die Vorgänge bis hinunter zur letzten Synapse akkurat beschreibt. Aber

[20] Ich spreche *nicht* von Intuitionen darüber, welche Sorte von Prozessen Geist hervorbringen kann, sondern, ‚wo' im Chinesischen Zimmer Geist zu finden sein könnte.

> genausogut kann man sich eine Computersimulation der Oxidation von Benzin in einem Automotor oder des Ablaufs von Verdauungsprozessen nach dem Verspeisen einer Pizza denken. Und die Simulation ist im Falle des Gehirns um nichts realer als im Falle des Autos oder des Magens: Wunder ausgeschlossen, können Sie kein Auto durch eine Computersimulation der Verbrennung von Benzin zum Laufen bringen und keine Pizza verdauen, indem Sie ein Programm laufen lassen, das die Verdauung simuliert. Entsprechend kann ganz offensichtlich auch eine Simulation von Denkvorgängen nicht die realen Effekte der Neurobiologie des Denkens erzeugen. (S. 43, rechts)

Searle spricht hier im gleichen Atemzug von einem „Computermodell mentaler Prozesse“ und der „Simulation von Denkvorgängen“ einerseits und von der „Computersimulation von Gehirnprozessen“ und der „Computersimulation der Wirkungsweise von Peptiden im Hypothalamus“ andererseits, als seien Peptide mentale Gegenstände. Damit vermischt er zwei Ebenen und erweckt den Anschein, es sei das Ziel der KI-Forschung, die *Duplikation* mentaler Phänomene durch ihre *Simulation* zu erreichen.[21] Zutreffender wäre m. E. die Beschreibung, daß mittels der Simulation der dem Geist *zugrundeliegenden* Prozesse eine Duplikation des Geistes selbst angestrebt wird.[22] Denken wir noch einmal an die oben eingeführte Idee der Hierarchie von Abstraktionsebenen zwischen der Hardware- und der mentalen Ebene, so können wir das auch anders ausdrücken: Durch die Simulation der Prozesse auf einer *niedrigen* Abstraktionsebene (eine sehr ‚feinkörnige‘ Simulation sozusagen, z. B. eine der neuronalen Ebene) sollen die Phänomene der *obersten* Abstraktionsebene reproduziert werden.

Natürlich hat Searle recht damit, daß auch die feinstkörnige Simulation eines Gehirns niemals ein Gehirn sein wird.[23] Die Frage ist nur, ob wirklich der Stoff, aus dem die Hirne sind, auch der einzige Stoff ist, aus dem Geist erwachsen kann. Darauf will Searle sich gar nicht versteifen. Er schließt nicht aus, daß mit geeigneten Programmen versehene Elektronenrechner tatsächlich Bewußtsein hervorbringen könnten, denn Siliziumchips könnten ja zufällig kausale Kräfte besitzen, die denen des menschlichen Gehirns äquivalent sind. Seine Behauptung ist nicht, daß Elektronenrechner niemals denken könnten, sondern daß Computer gleich welcher Bauart niemals *allein aufgrund*

[21] Unter einer „Simulation mentaler Vorgänge“ stelle ich mir etwas in der folgenden Art vor: Auf dem Gelände einer Spielzeugeisenbahn fahren zwei Züge aufeinander zu. Auf den ersten Waggon des einen ist geschrieben: „Schokolade“, auf den zweiten: „lecker“, auf den dritten: „also“ und auf den vierten: „kaufen“. Auf den Waggons des anderen Zuges steht: „Schokolade“, „macht fett“, „also“, „nicht“, „kaufen“. Die beiden Züge kollidieren, der zweite entgleist und der erste fährt in den Bahnhof ein. Das Ganze wäre eine Simulation eines Entscheidungsprozesses im Supermarkt. Die Waggons repräsentieren bestimmte Konzepte und Dinge in der Welt, die Züge sind Modelle von Gedanken und die Entscheidung fällt zugunsten der Schokolade. Eine solche Simulation wäre zu oberflächlich, um von Belang zu sein.

[22] Aber auch diese Beschreibung trifft natürlich nicht wirklich zu. Ziel der (‚starken‘) KI-Forschung ist es nicht, auf Biegen und Brechen Maschinen zum Denken zu bringen, sondern Maschinen zum Denken zu bringen *und* dabei das Denken zu *verstehen*. Angenommen, es wäre gelungen, die neurophysiologische Struktur eines menschlichen Gehirns vollständig zu analysieren und im Computer detailgetreu zu modellieren: Der einzige Gewinn bestünde darin, daß man an diesem ‚Gehirn‘ eventuell etwas ungehemmter experimentieren und herumpfuschen dürfte als an einem menschlichen. Aber weder könnte man durch die Leistung an sich Skeptikerinnen wie Searle überzeugen (s. „The Brain Simulator Reply“ in Searle 1980, S. 420 f), noch wüßte man dadurch mehr darüber, *wie* diese komplexen Prozesse Geist zustandebringen bzw. ausmachen. Die KI-Methode besteht vielmehr darin, Theorien darüber aufzustellen, Prozesse welcher Struktur Geist hervorbringen, diese Theorien in Programme umzusetzen, diese auf ihre Leistungen zu testen und sich dann zu überlegen, ob das Gehirn ähnlich arbeitet.

[23] Interessant wäre allerdings zu hören, wie Searle *erklärt*, warum er immer noch kein Chinesisch versteht, nachdem er das Chinesisch-Programm auswendiggelernt hat, da doch in diesem Fall alle notwendigen Ingredienzien beisammen zu sein scheinen: ein geeignetes Programm und ein geeignetes Gehirn.

ihrer Programmierung denken könnten. Es geht ihm nicht um die fehlenden kausalen Kräfte von *Computern*, sondern um die von *Programmen*.

Er scheint jedoch nicht sicher zu sein, ob Programme nun überhaupt keine kausalen Kräfte haben oder nur zu schwache. So schreibt er einerseits: „Symbole und Programme [sind] rein abstrakte Gebilde: Sie haben keinerlei intrinsische physikalische Eigenschaften, und sie lassen sich in jedem beliebigen physikalischen System darstellen. Somit haben auch Nullen und Einsen in ihrer Funktion als Symbole keinerlei intrinsische physikalische Eigenschaften und damit insbesondere auch keine kausalen Eigenschaften wie reale physikalische Objekte“ (S. 41, Mitte). Andererseits schreibt er aber auch: „Die einzige Kraft, die Symbolen als solchen innewohnt, ist die Fähigkeit, den nächsten Schritt eines Programms auszulösen, wenn der Computer läuft“ (S. 45, Mitte). Diese Fähigkeit hält er offenbar für irrelevant.

Selbstverständlich erwartet niemand, daß ein völlig losgelöst im leeren Raum schwebendes oder auf einen Berg von Papier gedrucktes KI-Programm irgendetwas bewirkt (Popper eventuell ausgenommen), geschweige denn Bewußtsein besitzt (vielleicht muß man es als körperlose Seele auffassen?). Es geht darum, was ein Programm bewirken kann, wenn es in irgendeinem Computer implementiert ist.

Daß es nicht darauf ankommt, aus was für einem Material dieser Computer besteht, hält Searle für das entscheidende Handicap jedes KI-Programmes, das für die Zusprechung von Geist kandidiert. Er gibt mehrere Beispiele an, aus was für unterschiedlichen Stoffen Computer bestehen können: „Man könnte im Prinzip einen Computer aus mit Drähten verbundenen alten Bierdosen bauen, der durch Windmühlen angetrieben wird; tatsächlich hat Joseph Weizenbaum vom Massachusetts Institute of Technology gezeigt, wie man einen (sehr langsamen) Heimcomputer aus Toilettenpapier und kleinen Steinchen basteln kann“ (S. 43, links). In Searle 1980 (S. 421, links) bringt er auch das Beispiel eines Computers (eigentlich eines Gehirnsimulators), der aus mit Ventilen versehenen Wasserleitungen besteht. Programme sind also nicht von Mikrochips abhängig, sondern können in einer Vielzahl von Maschinen mit einer ebensolchen Vielfalt kausaler Kräfte ablaufen. Die einzigen kausalen Kräfte, die einem Programm eigentümlich sind, sind die nicht an spezielle Materialien gebundenen Kräfte des Programms, Prozesse bestimmter *Struktur* ablaufen zu lassen, wenn es in Gang gesetzt wird.

Darüber, ob nun diese Kräfte schon zum Erzeugen von Geist ausreichen oder nicht, d.h. ob das Wesentliche an mentalen Vorgängen einzig in ihrer Struktur liegt oder auch in der physikalischen Beschaffenheit ihrer Hardwarebasis, wird man lange (und ich glaube, ergebnislos) streiten können. Ich denke, wer die starke KI-These vertritt, geht von zwei Grundannahmen aus: Erstens, daß mentale Prozesse auf physikalische Prozesse reduzierbar sind, und zweitens, daß für Geist nicht diese physikalischen Prozesse selbst relevant sind, sondern nur ihre Struktur.[24] Der ersten Hypothese hängt auch Searle an, gegen die zweite (u.a.) richtet sich sein Artikel. Er bringt jedoch m.E. kein überzeugendes Argument dafür vor, daß diese Hypothese falsch ist. So bleibt auch diese Frage meiner Meinung nach Glaubenssache.

Die Churchlands meinen, es sei eine empirische Frage (S. 54, links). Ich weiß nicht, wie sie auf diese Idee kommen. Mit empirischen Daten werden sie eine Skeptikerin wie Searle eben nicht umstimmen können. Denn was wäre ein empirischer Beleg dafür, daß irgendeine Maschine Bewußtsein besitzt? Bestenfalls kann sie den Turing-Test

[24] D.h. es gibt einen Grad der ‚Feinkörnigkeit‘ der Simulation, der für die Erzeugung von Geist hinreichend ist. Und wenn es nicht genügt, die Struktur der neuronalen Ebene zu reproduzieren, dann wird eben eine Ebene tiefer angesetzt. Und selbst wenn man das Verhalten jedes einzelnen Elementarteilchens im Gehirn simulieren müßte: Irgendwo hat dieser Abstieg ein Ende.

bestehen, und der ist eben immer noch kein *Nachweis* von Intentionalität (s. Fußnoten 4 und 12).

Abgesehen davon, daß Searle keine triftige Begründung dafür hat, daß die kausalen Kräfte des Gehirns zur Entstehung von Geist etwas Entscheidendes beitragen, bringt es ihm auch hinsichtlich seiner anderen Argumente nichts, auf diese Kräfte zu pochen. Ein dem vom Chinesischen Zimmer ähnliches Gedankenexperiment bringt z. B. Haugeland in seinem Kommentar zu Searle 1980:

> [Searle's Chinese Room] strategy will work as well against any specification of 'the right causal powers.' Instead of manipulating formal tokens according to the specifications of some computer program, the demon[25] will manipulate physical states or variables according to the specification of the 'right' causal interactions. Just to be concrete, imagine that the right ones are those powers that our neuron tips have to titillate one another with neurotransmitters. The green aliens can be intelligent, even though they're based on silicon chemistry, because their (silicon) neurons have the same power of intertitillation. Now imagine covering each of the neurons of a Chinese criminal with a thin coating, which has no effect, except that it is impervious to neurotransmitters. And imagine further that Searle's demon can see the problem, and comes to the rescue; he peers through the coating at each neural tip, determines which transmitter (if any) would have been emitted, and then massages the adjacent tips in a way that has the same effect as if they had received that transmitter. Basically, instead of replacing the c.p.u.,[26] the demon is replacing the neurotransmitters. By hypothesis, the victim's behavior is unchanged; in particular, she still acts as if she understood Chinese. Now, however, none of her neurons has the right causal powers – the demon has them,[27] and he still understands only English. Therefore, having the right causal powers (even while embedded in a system such that the exercise of these powers leads to 'intelligent' behavior) cannot be sufficient for understanding. Needless to say, a corresponding variation will work, whatever the relevant causal powers are. (Searle 1980, S. 432, links)

Interessant scheint mir vor allem Searles Interpretation dieses Gedankenexperiments:

> Her neurons still have the right causal powers; they just need some help from the demon. More generally if the stimulation[28] is at a low enough level to *reproduce* the causes and not merely *describe* them, the 'simulation' will reproduce the effects. (Searle 1980, S. 452 f)

Aber reproduziert der Dämon *wirklich* die Ursachen der neuronalen Erregungszustände, d. h. die Neurotransmitter? Nein, das tut er eben nicht: Er reproduziert ‚manuell' ihre physikalischen *Wirkungen*. Es laufen nicht die gleichen physikalischen Prozesse ab wie in einem unversehrten menschlichen Gehirn. Wenn ich also mit meinem Verständnis der ‚kausalen Kräfte' nicht sehr danebenliege, so sollte die Chinesin im Gedankenexperiment *kein* Chinesisch mehr verstehen. Oder sollte etwa diese ‚Simulation', weil sie auf einer „ausreichend niedrigen Ebene" ansetzt, doch eine *Duplikation* sein??? So offensichtlich wird Searle sich wohl nicht selbst widersprechen wollen.

[25] Haugeland bezeichnet das Wesen, das im Chinesischen Zimmer Symbole manipuliert, als „Searles Dämon", weil wir ihm im Rahmen des Gedankenexperiments übermenschliche Schnelligkeit bei der Symbolmanipulation unterstellen müssen. In Haugelands Gedankenexperiment kommt dann noch die Fähigkeit dazu, in mikroskopische Vorgänge einzugreifen.

[26] **C**entral **p**rocessing **u**nit, d. h. Zentralprozessor.

[27] Da bin ich allerdings anderer Ansicht.

[28] Das muß wohl „simulation" heißen.

Sein ‚formaler Beweis' läßt sich ebensoleicht gegen die tragende Rolle der kausalen Kräfte des Gehirns ummünzen:

Axiom 1: Neurophysiologische Prozesse sind rein physikalisch, also formal[29] (syntaktisch).

Axiom 2: Dem menschlichen Denken liegen geistige Inhalte (Semantik) zugrunde.

Axiom 3: Syntax an sich ist weder konstitutiv noch hinreichend für Semantik.

Folgerung: Neurophysiologische Prozesse sind weder konstitutiv noch hinreichend für Geist.

Mögliche Uminterpretationen formaler Systeme

Searle nennt noch ein weiteres Argument dagegen, daß das Chinesische Zimmer Chinesisch versteht. Er fragt, welchen Grund wir dafür hätten, die Vorgänge im Chinesischen Zimmer für spezifisch chinesische Geistestätigkeit zu halten. Die Chinesinnen, die mittels auf Kärtchen gedruckter chinesischer Symbole mit dem Zimmer ‚kommunizieren', könnten dazu geneigt sein, weil sie ohnehin schon eigene Interpretationen für diese Symbole haben und die ‚Antworten' des Zimmers zu diesen Interpretationen passen. Die Programmiererinnen, die das Regelbuch entworfen haben, nach dem Searle im Chinesischen Zimmer arbeitet, könnten dazu geneigt sein, weil das Verstehen chinesischer Sprache genau der Zweck ist, zu dem sie die Regeln (das Programm) entworfen haben. Beide Gruppen gehen also von vornherein davon aus, daß das Chinesische Zimmer sich mit chinesischer Sprache beschäftigt, und kommen daher kaum auf die Idee, die Möglichkeit anderer Interpretationen in Betracht zu ziehen. Searle schreibt:

> stellen Sie sich nun vor, daß mich, während ich im chinesischen Zimmer sitze und mit den chinesischen Symbolen herumhantiere, das Kombinieren der – mir völlig unverständlichen – Symbole zu langweilen beginnt. Ich beschließe daher, die Zeichen als Züge eines Schachspiels zu interpretieren. Welche Semantik erzeugt das System nun? Ist es eine chinesische Semantik, eine Schach-Semantik oder beides zugleich? Nehmen wir weiter an, es gäbe noch eine dritte Person, die durch ein Fenster hereinschaut und die Symbolmanipulationen als Börsenvorhersagen deutet. Und so fort. Es gibt keine Begrenzung für die Zahl der semantischen Interpretationen, die den Symbolen zugeordnet werden können, da – um es noch einmal zu sagen – diese rein formaler Natur sind; sie besitzen keine intrinsische Semantik. (S. 45 f)

Es mag sein, daß es „keine Begrenzung für die Zahl der semantischen Interpretationen" gibt, aber ich denke doch, daß die *Arten* akzeptabler Interpretationen sehr viel begrenzter sind, als Searle glaubt. Es gibt natürliche Interpretationen und unnatürliche. Die unnatürlichen würde wohl kaum jemand wirklich anerkennen wollen, und über die Verschiedenheit möglicher natürlicher Interpretationen brauchen wir uns m. E. keine Sorgen zu machen.

Wie könnte z. B. Searles ‚Schach-Semantik' beschaffen sein? Ein Extrembeispiel für eine *unnatürliche Interpretation* ist es, wenn eine bestimmte Schachpartie (sagen wir Sparov gegen Sparkassov am 30. 1. 1985 in Moskau) gegeben ist, und einfach festgelegt wird, daß *jede* beliebige Folge von Symbolmanipulationen im Chinesischen

[29] Vgl. Fußnote 11 und Hofstadters Kommentar zu Searle 1980 (S. 433, rechts, Mitte).

Zimmer *dieses* Schachspiel bezeichnen solle. Eine solche Interpretation ist unnatürlich, weil dabei die Bedeutung einer gegebenen Folge von Searles Handlungen im Zimmer nicht das Geringste damit zu tun hat, welche Handlungen in welcher Reihenfolge dies eigentlich waren. Interpretationen dieser Güteklasse sind billig zu haben. Man könnte sich z. B. irgendeinen Gullydeckel hernehmen und behaupten, in ihm sei das Alte Testament codiert. In einem Gullydeckel versteckt sich ebensogut das Alte Testament, wie sich Schachspiele im Wirken eines KI-Programms verstecken, das den Turing-Test bestehen kann.

Ich gebe ein Beispiel für eine sehr *natürliche Schach-Interpretation* von Searles Hantieren mit den Kärtchen: Angenommen, die Menge aller möglichen Handlungen, die Searle im Rahmen des Chinesischen Zimmers jeweils mit einem einzelnen Kärtchen durchführen könnte (z. B. „die oberste Karte von Stapel soundso nehmen" oder „diese Karte mit jener vergleichen" oder „von Stapel soundso das oberste Kärtchen wegnehmen, wenn es nicht das Krakel-Krakel-Symbol ist" – ich nenne sie mal *Prozessorhandlungen*, weil Searle sie in seiner Eigenschaft als Datenprozessor im Chinesischen Zimmer ausführt) – angenommen also, die Menge der Prozessorhandlungen ist in geeigneter Weise in mehrere Klassen unterteilt. Die Elemente der ersten Klasse interpretieren wir als Zeichen für die Farbe Schwarz, die aus der zweiten als Zeichen für die Farbe Weiß, die aus der dritten stehen für *Bauer*, die aus der vierten für *Turm*, etc., und die aus weiteren Klassen stehen jeweils für die Buchstaben und Zahlen, mit denen bestimmte Felder auf dem Schachbrett bezeichnet werden. Wir nehmen weiter an, daß diese Partition der Menge der Prozessorhandlungen geschickt zu den Abläufen im Chinesischen Zimmer und den Gegenständen der ‚Schachwelt' passend gewählt ist, so daß tatsächlich beispielsweise die Folge von Searles Handlungen, wenn er eine beliebige chinesische ‚Frage' verarbeitet, als Folge von Schachzügen deutbar ist (etwa „weißer Bauer E2–E4, schwarzer Bauer E7–E5, weißer Läufer G1–F3, . . . "). Ich bezweifle sehr, daß so ein Partition–Interpretation-Paar existiert, aber ich möchte es nicht völlig ausschließen. Es muß ja auch nicht unbedingt eine so einfache Interpretation sein, wie ich sie hier als Beispiel gewählt habe.

Je unnatürlicher die Interpretation, desto weniger hat das, was bei der Übersetzung herauskommt, mit den ursprünglichen ‚Zeichen' folgen zu tun; desto mehr erbt das, was bei der Übersetzung herauskommt, seine Information von den Übersetzungsregeln anstatt vom ursprünglichen ‚Text'. Vielleicht kann ich den Unterschied zwischen natürlichen und unnatürlichen Interpretationen mit Hilfe des *Isomorphismus*-Begriffs verdeutlichen. Eine Interpretation ist nichts anderes als eine Art Isomorphismus zwischen der Menge der Ausdrücke in einer ‚Sprache'[30] und der Menge der Ausdrücke in einer anderen.[31] Je *mehr* ein Isomorphismus von der relevanten Struktur bzw. Information der Ausdrücke der ersten Sprache beibehält, desto *natürlicher* ist er als Interpretation. Wer sehr unnatürliche Interpretationen zuläßt, könnte genausogut beispielsweise sagen, die Menge der natürlichen Zahlen sei ‚isomorph' zur Menge der komplexen Funktionen.[32] Schließlich sind beides Mengen. Ein solcher Billig-Isomorphismus ist jedoch überflüssig wie ein Kropf, weil er nichts mit der

[30] Mit *Sprache* meine ich dabei nicht nur natürliche Sprachen wie Suaheli und Deutsch. Ich verwende den Begriff so weit, daß im Prinzip alles, was als sauber interpretierbare Sammlung von Zeichen deutbar ist, ein Ausdruck in einer Sprache sein kann. (Was zugegebenermaßen nur eine recht schwammige Definition ist.)

[31] Oder eine Abbildung, die Ausdrücken in einer Sprache Phänomene in der Welt zuordnet (d. h. *Semantik*: Worten werden Dinge zugeordnet, Sätzen Sachverhalte (oder Wahrheitswerte), etc.). Da wir aber, um eine solche Abbildung anzugeben, diese Phänomene wiederum in irgendeiner Sprache benennen müssen, ist der Unterschied in diesem Zusammenhang nicht schwerwiegend.

[32] Das wäre natürlich kein Isomorphismus im strengen mathematischen Sinne.

Struktur der beiden Mengen zu tun hat; weil er nichts über sie aussagt.

Brauchbare Isomorphismen hingegen sagen uns, daß zwei Mengen einander sehr (‚zum Verwechseln') ähnlich sind, und darüberhinaus, welche Bestandteile der einen solchen der anderen entsprechen, so daß wir u. U. durch Betrachtung der einen Menge Neues über die andere erfahren können. Besäße ich eine gute Interpretation von Sätzen in Suaheli als deutsche Sätze, so könnte ich Leute verstehen, die mich auf Suaheli ansprechen, und dann aufgrund dieses Verständnisses handeln, etwa wie jemand gehandelt haben würde, der Suaheli als Muttersprache spricht.

Ich kann keine exakte Trennlinie zwischen natürlichen und unnatürlichen Interpretationen ziehen. Das ist jedoch nicht weiter schlimm, da wir auf die Betrachtung von Grenzfällen verzichten und uns mit den klaren Fällen begnügen können.

Ich bezweifle, daß Searle sehr unnatürliche Interpretationen im Sinn hatte. Hat Searle hingegen an eine natürlichere Interpretation gedacht, so frage ich mich, woher er den Glauben nimmt, eine *brauchbare* Schach-Semantik könne existieren. Ich habe geschrieben, ich würde im Extremfall zugestehen, daß eine einigermaßen natürliche Interpretation existiert, unter der die Folge von Searles Prozessorhandlungen, während er eine chinesische ‚Frage' verarbeitet, stets als Folge von Schachzügen deutbar ist. Aber ich bin mir sicher, daß keine solche Interpretation es ermöglicht, beliebige solche Folgen von Prozessorhandlungen als *gültige* Schachpartien zu deuten. KI-Programme haben einfach zuwenig mit Schach zu tun, als daß sie in halbwegs natürlicher Weise als sinnvolle Folge regelkonformer Schachzüge interpretiert werden könnten. Searle müßte also seine Langeweile weiterhin ertragen, es sei denn, er fände Gefallen an Nonsense-Schach. Ebenso bin ich überzeugt, daß jemand mit einer einigermaßen natürlichen ‚Börsen-Interpretation' der Vorgänge im Chinesischen Zimmer an der Wall Street unweigerlich Schiffbruch erleiden würde. Es gibt einfach keine natürliche Isomorphie zwischen KI-Datenverarbeitung und Schachspielen bzw. Börsenvorhersagen.

Der Glaube, derlei wäre möglich, rührt vielleicht teils von einer Selbsttäuschung über die Größe und die Komplexität der Struktur eines Turing-Test-tauglichen KI-Programmes her und teils von dem Irrtum, es genüge, Uminterpretationen jeweils für sehr kleine Bestandteile formaler Systeme zu finden. So mag es zwar richtig sein, daß die Differentialgleichung, die die Schwingungen von Spiralfedern beschreibt, nach geeigneter Uminterpretation der enthaltenen Parameter und Variablen genausogut als Beschreibung der Schwankungen der Stromstärke in bestimmten Stromkreisen interpretierbar ist (vgl. Carrier 1990, S. 9). Aber was ist, wenn man die formale Struktur der gesamten klassischen Mechanik nimmt und umzuinterpretieren versucht? Die Physikerinnen wären wahrscheinlich glücklich, wenn sie statt Mechanik *und* Elektrizitätslehre *nur noch* Mechanik treiben und die Ergebnisse dann mit Hilfe eines solchen Isomorphismus in die Termini der Elektrizitätslehre übersetzen müßten (oder andersherum). Ohne etwas von ihnen zu verstehen, bezweifle ich jedoch sehr, daß die beiden Theorien isomorph sind. Um wieviel schwieriger muß es sein, eine natürliche (und von der ‚kanonischen' verschiedene) Interpretation der Funktionsweise eines KI-Programmes zu finden, das den Turing-Test bestehen kann. Ein solches Programm müßte ja an Komplexität dem menschlichen Gehirn gleichkommen; darin müßte eine ganze ‚*Welt*' repräsentiert sein.[33] Und wohlgemerkt: Das Problem ist *nicht*, die *Vielzahl* mentaler Gegenstände in einem anderen Bereich wiederzufinden; in der Mathematik z. B. sind Mengen jeder gewünschten Größe bequem erhältlich. Das Problem

[33] Nicht in dem Sinne, daß ein solches KI-Programm allwissend sein müßte, sondern in dem, daß es annähernd genausoviel über die Welt wissen und glauben müßte wie ein Mensch. Eine ‚subjektive Welt' sozusagen.

besteht vielmehr darin, daß diese mentalen Gegenstände in sehr komplexer Weise miteinander *verknüpft* sind (z.B. durch logische Zusammenhänge und psychologische Assoziationen); für diese *Struktur* gilt es ein anderes Modell zu finden!

Für meinen Geschmack ist damit die Auswahl schon sehr stark beschränkt. Aber für alle Fälle setze ich noch einen drauf. Wir können uns das Chinesische Zimmer um einen Satz von Transducern (Fernsehkameras, Mikrofone, Thermometer, eventuell chemische Rezeptoren als Ersatz für Geruchs- und Geschmackssinn, etc.) erweitert vorstellen. Die optimale Ergänzung wäre ein Roboterkörper mit einem möglichst großen Teil der menschlichen Wahrnehmungs- und Handlungsfähigkeiten. Dieser könnte über Funk in Verbindung mit einem Computer stehen, der Kärtchen mit chinesischen Symbolen drucken und ‚lesen' kann und solche an Searle schickt und von ihm empfängt. Die Chinesinnen, deren ‚Kommunikation' mit dem Chinesischen Zimmer bislang darin bestand, mit Searle Stapel von mit chinesischen Symbolen bedruckten Kärtchen auszutauschen, könnten sich stattdessen nun direkt mit dem Roboter ‚unterhalten'. Die ‚Sinnesorgane' des Roboters digitalisieren alle hereinströmende Information und übermitteln sie dem an das Chinesische Zimmer angeschlossenen Computer. Dieser codiert die Informationen in geeigneter Weise in Folgen chinesischer Symbole um,[34] druckt diese auf Kärtchen und läßt sie Searle ins Zimmer plumpsen. Searle verarbeitet die Kärtchen (unter Verwendung eines neuen, der neuen ‚Versuchsanordnung' entsprechenden Regelbuches) und schiebt seine ‚Antwort' in den Rachen des Lesegeräts des Computers. Die ‚Antwort' geht den ganzen Weg zurück und wird schließlich vom Roboter in Handlungen (Bewegungen, Äußerungen) umgesetzt.

Wir können annehmen, daß das neue Programm ebenso geschickt mit dem Roboter umgeht, wie zuvor das alte in chinesischer Schrift korrespondierte, und weiterhin, daß Searle seine Symbolmanipulationen so ‚dämonisch' schnell durchführt, daß die Reaktionen des Roboters nicht langsamer als die eines Menschen sind. Searle gesteht (wenn auch in etwas anderem Zusammenhang) seinen Kritikerinnen zu, sein Gedankenexperiment in dieser Art abzuwandeln („The Robot Reply" in Searle 1980, S. 420, rechts). Ich glaube nicht, daß diese Erweiterung für die Intentionalität des Programmes notwendig ist,[35] aber wir engen damit den Spielraum für mögliche natürliche Interpretationen der Vorgänge im Chinesischen Zimmer weiter ein.

Denn nun muß eine solche Interpretation nicht nur die ‚*interne Struktur*' dieser Vorgänge respektieren, sondern zusätzlich ihr *Verhältnis zur Außenwelt*. D.h. wenn bestimmte Phänomene im Chinesischen Zimmer dann und nur dann auftreten, wenn der Roboter eine Taube ‚wahrnimmt', nicht aber, wenn er einen Bratapfel oder eine

[34] Das würden allerdings kaum chinesische *Sätze* sein, die beschreiben, was die Kamera ‚sieht'! Die Sehnerven von Deutschen telegraphieren schließlich auch keine deutschen Sätze an deren Gehirne.

[35] Ich halte es z.B. für plausibel, daß ein Mensch, der bei einem Unfall dergestalt verletzt wird, daß er keine Sinneswahrnehmungen und keine bewußte Kontrolle über seinen Körper mehr hat, oder ein künstlich am Leben gehaltenes, isoliertes Gehirn durchaus noch Geist besitzen können. (Allerdings dürfte dieser Zustand der geistigen Gesundheit eher abträglich sein.) Diesem Zustand des Abgeschnittenseins vom eigenen Körper gar nicht so unähnlich ist wohl der, in den man durch die sensorische Deprivation in einem sogenannten ‚Isolationstank' gerät. – Natürlich würde ein Mensch, der so *geboren* wird und bleibt, niemals Geist *erlangen*. Aber KI-Programme werden ja üblicherweise nicht als unwissende, unerfahrene, aber lernfähige ‚Baby-Programme' hergestellt (auch wenn das vielleicht eine bequeme Strategie wäre, Maschinen zum Denken zu bringen), sondern in einem Zustand der ‚Reife'. Um diejenigen Philosophinnen zum Schweigen zu bringen, die meinen, Geist könne nur haben, wer vorher Erfahrungen mit seiner Umwelt gemacht hat, könnte man eventuell ein Programm mit ‚künstlichen Erinnerungen' versehen, d.h. es so gestalten, wie ein Reifungsprozeß vom Baby-Programm zur ausgewachsenen künstlichen Intelligenz es geformt haben würde. Die Künstlichkeit dieser ‚Vergangenheit' würde für das Programm keinen Unterschied bedeuten.

Schneewehe ‚wahrnimmt', dann drängt es sich geradezu auf, diese Phänomene als interne Repräsentation des Begriffs *Taube* zu interpretieren. Ein mathematisches Analogon zu dieser Restriktion der möglichen natürlichen Interpretationen ist das Folgende: Die Menge der komplexen Zahlen als *reeller* Vektorraum ist durchaus *nicht* isomorph[36] zur Menge der komplexen Zahlen als *komplexer* Vektorraum, obwohl die Menge der Vektoren (d.h. die additive Gruppe der komplexen Zahlen) in beiden Fällen dieselbe, insbesondere isomorph ist.

Die nun übrigbleibenden möglichen natürlichen Uminterpretationen von Searles Prozessorhandlungen dürften m.E. nur noch von der Art sein, wie sie Quine in *Theories and Things* beschreibt (dort bezogen auf Übersetzungen aus (radikal) fremden Sprachen in bekannte). Ein Beispiel für eine solche Uminterpretation funktioniert wie folgt: Namen von Gegenständen werden nicht mehr als diese Gegenstände *selbst* bezeichnend interpretiert. Stattdessen wird der Name eines Gegenstandes aufgefaßt als Bezeichnung der *Menge der Punkte* im vierdimensionalen Raum-Zeit-Kontinuum, die von diesem Gegenstand im Laufe seines Daseins eingenommen werden. Angenommen, jemand besitzt einen Hund namens Fido, so bezeichnet der Name *Fido* unter der neuen Interpretation also nicht mehr ein Tier aus Fleisch und Blut, sondern einen ‚Raum-Zeit-Schlauch', der sich von der Geburt dieses Tieres bis zu seinem Tod durch die Zeit erstreckt.

Die Interpretation soll natürlich sein, daher werden die Prädikate bzw. die Namen von Begriffen (z.B. wenn wir aus dem Englischen übersetzen: *dog*, *blue*) entsprechend umgedeutet. *Fido is a dog* als *Der und der Raum-Zeit-Schlauch ist ein Hund* zu übersetzen, wäre Unsinn. Punktmengen sind keine Hunde. Also wird das Prädikat *dog* umgedeutet als *Raum-Zeit-Stelle eines Hundes*. Allgemeiner wird ein Prädikat *P* jeweils uminterpretiert als *Raum-Zeit-Stelle eines P*. So bleibt die Bedeutung *ganzer Sätze* jeweils die gleiche, und die neue Interpretation macht Sinn.

Die Frage, worauf in der Welt sich Searles Prozessorhandlungen beziehen, können wir nun ähnlich gut beantworten wie die Frage, worauf sich Äußerungen in chinesischer Sprache beziehen: In beiden Fällen gibt es einen gewissen Interpretationsspielraum, der aber zu gering ist, als daß wir uns darüber Sorgen machen müßten. Ebenso, wie wir Chinesinnen unter den verschiedenen sinnvollen Übersetzungen sehr gut verstehen, können wir sicher sein, daß Searle im Chinesischen Zimmer nicht etwa unwissentlich Schachspiele durchexerziert oder Börsenkurse prophezeit. Möglich, daß das, was wir für eine interne Repräsentation des Begriffes *Taube* halten möchten, in Wirklichkeit etwas anderes repräsentiert, aber eine ähnliche Unsicherheit hält uns im Falle der Chinesin auch nicht davon ab, ihr ‚spezifisch chinesische Gedankeninhalte' zuzusprechen.

Material-Mystizismus

Am Ende seines Artikels spekuliert Searle noch etwas boshaft, die wissenschaftliche Verirrung, an die starke KI-These zu glauben, rühre wohl von einer paradoxen Mischung aus Dualismus und Behaviorismus her. Entsprechend möchte ich mir zum Schluß das Vergnügen des Versuches gönnen, *Searles* Gedankengang zu rekonstruieren.

Ich vermute, am Anfang von Searles Überlegungen stand das Mißverständnis, in der KI-Forschung solle *Computern das Denken beigebracht* werden, indem man ihnen Programme zur Ausführung gibt. (Zugegebenermaßen ist die Sprechweise vom

[36] Diesmal im strengen mathematischen Sinne.

denkenden Computer von verführerischer Bequemlichkeit, und ich habe sie auch selbst ein paarmal gebraucht.) Die sauberere Beschreibung ist jedoch m. E., daß versucht wird, *Programme herzustellen* (zu schreiben), *die denken*, wenn man sie in Computern implementiert.

Einmal diesem Mißverständnis erlegen, konnte Searle natürlich nur noch schwer an die Möglichkeit glauben, dieses Forschungsvorhaben könne erfolgreich sein. Wie sollte denn der Computer denken können, indem er nach irgendwelchen vertrackten Vorschriften elektronisch herumfuhrwerkt? Der Computer mochte seine Binärzahlen herumschieben, so viel er wollte, dadurch würde er sie auch nicht besser verstehen; geschweige denn ein Gefühl für das Unbehagen der postindustriellen Gesellschaft im Spätkapitalismus kriegen. Genauso könnte Searle chinesische Symbole manipulieren, bis er schwarz würde, ohne deswegen jemals ein Wörtchen Chinesisch zu verstehen. Der Fehler bei seinem Gedankenexperiment besteht wiederum darin, daß Searle sich mit dem Computer identifiziert anstatt mit dem Programm. Die Rolle von jenem ist natürlich auch leichter zu übernehmen: Was der Computer tut, kann Searle auch (wenngleich nicht so schnell), aber *das Programm sein* kann er nicht, nicht einmal, indem er es auswendig lernt.

Computer sind doof, das bestreitet niemand. Alles, was Computer haben, sind die Regeln, nach denen sie arbeiten; ist *Syntax*, und davon kann man ihnen einflößen, soviel man will: Sie werden trotzdem nicht schlauer; sie gelangen nie zur *Semantik*. Darauf beruht Searles ‚formaler Beweis'. (Der in Wirklichkeit nicht sehr formal ist, obwohl die Axiome sauber durchnumeriert sind.) D. h. er beruht auf dem Irrtum, die inneren Repräsentationen und die mentalen Gegenstände auf einer viel zu niedrigen Ebene zu vermuten.

Aber Neuronen sind doch auch doof? Wie also kommt es, daß das Gehirn Gedanken zustandebringt, der Computer aber nicht? Auf der Suche nach einer Erklärung versucht Searle, noch materialistischer als die Materialistinnen zu sein, und verfällt dabei in eine Art *Material-Mystizismus*. Nicht nur kreist die Erde um die Sonne und stammt der Mensch von Affen ab, nein, auch am menschlichen Geist ist nichts Wunderbares; er ist genauso erdverbunden, *materialverhaftet*, wie die Verdauung von Pizza und die Verbrennung von Benzin. Das Material macht's. Gehirne aus Eiweiß verstehen Sprache, Gehirne aus Wasserleitungen blubbern bloß syntaktisch dahin. Eine kuriose Mischung von Materialismus und Anthropozentrismus ist das, so scheint mir.

Literatur

Carrier, Martin. 1990. „On the Disunity of Science, or: Why Psychology Is not a Branch of Physics". *Konstanzer Berichte zur Logik und Wissenschaftstheorie* 6.

Churchland, Paul M., und Patricia Smith Churchland. 1990a. „Could a Machine Think?" *Scientific American*, 1:26–31.

———, und ———. 1990b. „Ist eine denkende Maschine möglich?" *Spektrum der Wissenschaft*, 3:47–54. [Die deutsche Übersetzung des vorigen Artikels.]

Hofstadter, Douglas R. 1980. *Gödel, Escher, Bach: an Eternal Golden Braid*. New York: Vintage. [Ein wunderschönes Buch, von dem sich die meisten Autorinnen wissenschaftlicher Bücher in punkto Didaktik manch dicke Scheibe abschneiden könnten.]

———. 1986. *Metamagical Themas: Questing for the Essence of Mind and Pattern*. New York: Bantam. [Enthält u. a. einen sehr guten Dialog zu der Frage, wie der freie Wille des Geistes mit Determinismus im Gehirn zusammengehen kann: „Who Shoves Whom around inside the Careenium? or, What Is the Meaning of the Word 'I'?", S. 604–627.]

———, und Daniel C. Dennett. 1981. *The Mind's I: Fantasies and Reflections on Self and Soul*. Brighton: Harvester Press. [Enthält auch Searle 1980 und kommentiert ihn, aber es gibt bessere Kommentare dazu.]

Quine, Willard Van Orman. 1981. *Theories and Things*. Cambridge (Mass.).

Searle, John R. 1980. „Minds, Brains and Programs". *Behavioral and Brain Sciences*, 3:417–456. [Der ursprüngliche Artikel steht auf den Seiten 417–424. Auf den Seiten 424–450 befinden sich die Kommentare von 28 anderen Wissenschaftlerinnen. Einige von diesen Kommentaren finde ich interessant (z. B. Haugeland (S. 432 f), Pylyshyn (S. 442 ff), Wilensky (S. 449 f)), einige spaßig (z. B. Minsky (S. 439 f)) und manche blöd (z. B. Puccetti (S. 441 f), der uns zeigt, daß Schachprogramme Schach nicht wirklich verstehen). Auf den Seiten 450–456 geht Searle dann wiederum auf seine Kritikerinnen ein.]

———. 1990a. „Is the Brain's Mind a Computer Program?" *Scientific American*, 1:20–25. [Gewissermaßen eine verbesserte und erweiterte Neuauflage von Searle 1980.]

———. 1990b. „Ist der menschliche Geist ein Computerprogramm?" *Spektrum der Wissenschaft*, 3:40–47. [Die deutsche Übersetzung des vorigen Artikels.]

Turing, Alan M. 1950. „Computing Machinery and Intelligence". *Mind*, 59:433–460. [Turing stellt das Imitationsspiel vor, das, wenn man es mit KI-Programmen spielt, heute Turing-Test genannt wird.]