

Christian Nimtz 2012 Universität Bielefeld	to be published: <i>Christian Nimtz: Das Chinesische Zimmer, in: Rolf W. Puster (Hg.): Klassische Argumentationen der Philosophie, Paderborn: mentis 2012.</i>
---	--

Das Chinesische Zimmer

Christian Nimtz

cnimtz@uni-bielefeld.de

I.

Das Chinesische Zimmer (*Chinese room*) ist ein weit über die Philosophie hinaus diskutiertes Gedankenexperiment des amerikanischen Philosophen John Searle (*1932). Searle hat das Gedankenexperiment zuerst in seinem Artikel „Minds, Brains, and Programs“ aus dem Jahr 1980 vorgestellt und in einer Reihe späterer Veröffentlichungen verteidigt.¹ Mit Hilfe des Experiments will Searle zunächst einmal die sogenannte Computertheorie des Geistes widerlegen und das Forschungsprogramm der Künstlichen Intelligenz als Irrweg entlarven. So erstaunt es wenig, dass Searles Überlegung ein ungemein breites – und durchaus nicht nur kritisches – Echo aus den Reihen der Kognitionswissenschaftler erfahren hat.² Der Sache nach will Searle mit seinem Szenario des Chinesischen Zimmers aber nicht nur ein bestimmtes Forschungsprogramm treffen. Er will vielmehr das seit Mitte der 1960er Jahre die Theoriebildung in Philosophie und Psychologie prägende Paradigma des Funktionalismus als falsch erweisen. Ob ein Wesen denkt, hängt dem Funktionalismus zufolge nicht davon ab, woraus das Wesen (oder sein Gehirn) besteht. Denkende Wesen können aus ganz unterschiedlichem Material bestehen; neben denkenden Wesen auf Kohlenstoffbasis wie uns kann es auch z.B. denkende Wesen auf Basis von Silizium geben. Ob ein Wesen denkt, hängt dem Funktionalismus zufolge vielmehr davon ab, ob es in ihm (bzw. in seinem Gehirn) Zustände gibt, die die bestimmte kausale Rollen spielen, d.h. die mit anderen Zuständen, Eingaben

¹ Vgl. insbesondere die Vorlesungen „Geist, Hirn und Wissenschaft“ (Searle 1986), Searle 1990, Searle 1993, Kap. 9 und Searle 2002.

² Vgl. die Artikel in Preston/Bishop 2002, die dort im Anhang abgedruckte Bibliographie und die Literaturliste in Cole 2009. Siehe auch die betreffende Sektion in David Chalmers' online-Bibliographie: <http://consc.net/mindpapers/6.1c>.

und Ausgaben auf spezifische Weise kausal verknüpft sind. Diese funktionalistische Idee hält Searle für grundfalsch; ihr will er die Grundlage entziehen. Dazu beruft sich Searle nicht auf empirische Daten. Er will die Computertheorie des Geistes und den Funktionalismus gleichsam vom Lehnstuhl aus durch ein Gedankenexperiment widerlegen. Daraus erklärt sich die generelle philosophische Brisanz seiner Überlegung.

II.

1. *Die Computertheorie des Geistes*³

Um die Computertheorie des Geistes und das mit ihr verschränkte Forschungsprogramm der Künstlichen Intelligenz verstehen zu können, müssen wir zunächst einmal die Begriffe *Syntax*, *Semantik*, *Symbol* und *Symbolverarbeitung* klären. Am besten beginnen wir mit einem Beispiel. Denken Sie an das Wort „Go“. Auf der einen Seite hat dieses Zeichen formale Eigenschaften – es besteht aus den Zeichen „G“ und „o“ und kann in wohlgebildeten Sätzen des Deutschen wie „Go ist ein Spiel“ nur auf bestimmte Weise vorkommen. Diese Eigenschaften bezeichnet man als die *syntaktischen* Eigenschaften des Zeichens, und man fasst die Gesamtheit der syntaktischen Eigenschaften einer Sprache oder eines Zeichensystems als deren *Syntax* zusammen. (Zugleich versteht man unter ‚Syntax‘ die allgemeine Theorie der formalen Eigenschaften von Sprachen bzw. Zeichensystemen.) Auf der anderen Seite hat das Zeichen „Go“ eine bestimmte Bedeutung: Es bezeichnet ein japanisches Brettspiel. Die zur Bedeutung eines Zeichens gehörenden Eigenschaften bezeichnet man als die *semantischen* Eigenschaften des Zeichens, und man fasst die Gesamtheit der semantischen Eigenschaften einer Sprache oder eines Zeichensystems als deren *Semantik* zusammen. (Zugleich versteht man unter ‚Semantik‘ die allgemeine Theorie der semantischen Eigenschaften von Sprachen bzw. Zeichensystemen.)

Was ist nun ein Symbol? Im Alltag verstehen wir darunter ein bestimmtes Zeichen *mit* seiner Bedeutung. Das ist auch in der Computertheorie des Geistes nicht anders. Symbole können semantische Eigenschaften haben – etwas *bezeichnen* oder über einen *semantischen Gehalt* verfügen – oder, wie man sagt, ‚interpretiert‘ sein. Allerdings spielen dieser Theorie zufolge bei der Verarbeitung von Symbolen deren syntaktische Eigenschaften eine Rol-

³ Zur ersten Einführung vgl. Ravenscroft 2008, Kap. 6 und Beckermann 2008, Kap. 6. Eine zugängliche Einführung in die KI bietet nach wie vor Haugeland 1987. Eine solide Orientierung über die KI bieten Russell/Norvig 2002.

le. Entsprechend wird unter *Symbolverarbeitung* die regelgeleitete Überführung von Zeichenketten in Zeichenketten allein aufgrund ihrer syntaktischen Eigenschaften begriffen. Symbolverarbeitung in diesem Sinne ist heute allgegenwärtig. Letztlich ist ein jeder Computer nichts anderes als eine symbolverarbeitende Maschine – d.h. ein System, das Zeichenketten nach gegebenen Regeln rein mechanisch anhand ihrer syntaktischen Eigenschaften manipuliert, dabei von internen Zuständen in andere interne Zustände übergeht und auf diese Weise zu Eingaben bestimmte Ausgaben erzeugt.

Als mechanische symbolverarbeitende Systeme sind Computer rein syntaktisch operierende Maschinen. Trotzdem reicht ihre Bedeutsamkeit über die syntaktische Verarbeitung hinaus. Auf der einen Seite hat der englische Beweistheoretiker Alan Turing (1912–1954) plausibel gemacht, dass sich im Prinzip jede beliebige durch einen Algorithmus – d.i. eine Lösungsvorschrift in endlich vielen Schritten – lösbare Rechenaufgabe mittels mechanischer Symbolverarbeitung lösen lässt. Wir können demnach einen Computer für jede solche Aufgabe so programmieren, dass er für Zahlzeichen als Eingabe durch rein syntaktische Manipulation dasjenige Zahlzeichen ausgibt, das das korrekte Resultat ausdrückt. Auf der anderen Seite können wir einen Rechner so einrichten, dass er rein durch Symbolverarbeitung zu Sätzen als Eingabe die korrekten logischen Folgerungen ausgibt und z.B. zur Eingabe von „Alle Menschen sind sterblich“ und „Sokrates ist ein Mensch“ die Ausgabe „Sokrates ist sterblich“ produziert. Zumindest in einigen Bereichen lassen sich symbolverarbeitende Maschinen mithin so einrichten, dass ihre rein syntaktische Verarbeitung von Zeichen mit den semantischen Beziehungen zwischen den Zeichen konform geht – gerade so, wie dies bei den erwähnten Rechenaufgaben und der angeführten Folgerung der Fall ist. Kurz gesagt kann rein syntaktische Symbolverarbeitung *gehaltkonform* sein.

Die Einsicht, dass rein syntaktische Symbolverarbeitung gehaltkonform sein kann, bildet den Grundstein des Forschungsprogramms der sogenannten *Künstlichen Intelligenz*. Dieses Programm verbindet eine computerwissenschaftliche und eine kognitionswissenschaftliche Forschungsrichtung. Dem Programm der Künstlichen Intelligenz verpflichtete Computerwissenschaftler wollen „Maschinen entwickeln, die Funktionen ausführen, für deren Ausführung Menschen Intelligenz benötigen“ (Kurzweil 1990, 11#). Dies hat sich als schwieriger herausgestellt, als anfänglich gedacht. Aber heutzutage erfüllen viele Computersysteme diesen Anspruch. Sie steuern kleine Staubsaugroboter unfallfrei durch unsere Wohnungen, überwachen

Piloten bei Start und Landung und organisieren den logistisch optimalen Transport von Material und Personal. Einige Computersysteme können sogar mit Großmeistern im Schach mithalten oder Autos selbstständig durch unbekanntes Terrain lenken. (Merke: Oft sind alltägliche und von vielen dummen Menschen gemeisterte Aufgaben für Computersysteme besonders herausfordernd.)

Die zweite im Programm der Künstlichen Intelligenz vertretene Forschungsrichtung will nicht leistungsfähigere Computer bauen, sondern den menschlichen Geist verstehen. Dem Programm der Künstlichen Intelligenz verpflichtete Kognitionswissenschaftler und Philosophen verfechten eine als *Computertheorie des Geistes* bekannte Theorie über die Natur des Mentalen. Die Grundidee dieser insbesondere von Jerry Fodor in der Philosophie des Geistes einflussreich entwickelten Theorie lautet so:

*Der menschliche Geist ist ein System, in dem semantisch gehaltvolle Symbole auf gehaltkonforme Weise syntaktisch verarbeitet werden.*⁴

Ein Verfechter der Computertheorie des Geistes legt sich auf eine Theorie mentaler Zustände fest: Die Wünsche, Überzeugungen, Absichten etc. eines Denkers bestehen darin, dass in seinem Geist an funktional relevanter Stelle Symbole mit einem betreffenden semantischen Gehalt vorkommen. Mein Haben der Überzeugung, die Erde kreise um die Sonne, besteht demnach darin, dass in meinem Überzeugungsspeicher ein Symbol mit dem semantischen Gehalt, *dass die Erde um die Sonne kreist*, abgelegt ist. Zugleich verpflichtet er sich auf eine Theorie über Denkprozesse: Denkprozesse bestehen in der regelgeleiteten Manipulation von Symbolen allein aufgrund ihrer syntaktischen Eigenschaften. Wenn ich aus der gerade beschriebenen Überzeugung die Ansicht ableite, dass etwas um die Sonne kreist, dann besteht der betreffende Prozess in Symbolverarbeitung, die zu einem Symbol mit dem semantischen Gehalt, *dass etwas um die Sonne kreist*, in meinem Überzeugungsspeicher führt.

Die Computertheorie des Geistes ist eine erklärungs mächtige und entsprechend populäre Theorie. Erstens bietet sie, wie gerade beschrieben, neben einer Theorie mentaler Zustände auch eine ausgeführte Theorie mentaler Prozesse und damit eine Theorie des Denkens. Aufgrund des zugrundegelegten Modells der Symbolverarbeitung kann diese Theorie dazu zentralen Eigenschaften menschlichen Denkens wie dessen Produktivität – wir

⁴ Vgl. Fodor 1987, Kap. 1 und Appendix sowie Fodor 2008; für eine Übersicht vgl. Horst 2009.

können nie zuvor gedachte Gedanken denken – und Systematizität – wer denken kann, dass die Erde um die Sonne kreist, kann auch denken, die Sonne kreise um die Erde – Rechnung tragen. Zweitens können Anhänger der Computertheorie eine einleuchtende Antwort auf die Frage nach dem Verhältnis von Geist und Gehirn geben. Substanzdualisten wie René Descartes betrachten unseren Geist als eine immaterielle *res cogitans* und weisen dem strikt von diesem verschiedenen Gehirn allein die Rolle zu, kausaler Angriffspunkt für das Einwirken des Geistes auf den Körper zu sein. Eine solche Position kann nicht erklären, warum wir über ein so komplexes Gehirn verfügen, wo doch der Geist alle kognitiven Leistungen vollbringt.⁵ Identitätstheoretiker dagegen identifizieren mentale Eigenschaften mit physischen Eigenschaften und setzen – arg vereinfacht – beispielsweise die Wut einer Person mit einer spezifischen Erregung ihrer Amygdala gleich. Eine solche Position kann nicht erklären, wie Lebewesen mit einer ganz anderen Neurophysiologie dieselben mentalen Zustände haben können wie wir.

Im Einklang mit dem anfangs erwähnten funktionalistischen Paradigma, das mentale Zustände über kausale Rollen erklärt, schlägt die Computertheorie des Geistes eine dritte Antwort vor: Der Geist einer Person besteht in Zuständen und Prozessen, die durch die physischen Strukturen seines Gehirns *realisiert* sind. Unser Geist verhält sich demnach zu unserem Gehirn in etwa wie ein Programm (Software) zu dem Computer (Hardware), auf dem es abläuft. Einerseits ist unser Geist also von unserem Gehirn abhängig, denn ohne eine entsprechende physische Basis kann plausiblerweise keine mechanische Symbolverarbeitung stattfinden. Damit kann die Computertheorie erklären, warum wir ein komplexes Gehirn benötigen: Es dient als physische Grundlage für das komplexe Symbolsystem aus Zuständen und Verarbeitungsprozessen, das unseren Geist ausmacht. Andererseits ist unser Geist von unserem Gehirn unabhängig, denn dasselbe Symbolsystem kann auch auf einer anderen physischen Grundlage realisiert sein. Die Computertheorie kann also erklären, wie Lebewesen mit einer ganz anderen Neurophysiologie dieselben mentalen Zustände haben können wie wir. In diesen ist dasselbe Symbolsystem in einem anderen physischen Medium – sprich: einem anderen Gehirn – realisiert. Diese Idee der *multiplen Realisierbarkeit* mentaler Zustände erklärt nicht nur mentale Gleichheiten über Spezies-schranken hinweg. Sie leistet auch der Idee Vorschub, es könne Computersysteme geben, die im Wortsinne denken. Denn wenn die kognitiven Ver-

⁵ Vgl. Beckermann 2008, Kap. 2 und 3.

mögen von uns Menschen auf Symbolverarbeitung beruhen und dasselbe Symbolsystem in verschiedenen physischen Grundlagen realisiert sein kann, warum sollen sich Symbolverarbeitungsprozesse wie die, die menschlichem intelligenten Denken zugrunde liegen, nicht auch in Computern oder Robotern realisieren lassen?

Drittens schließlich kann die Computertheorie des Geistes erklären, wie es in einer fundamental physischen Welt wie der unsrigen denkende Wesen geben kann. Der Computertheorie zufolge ist der menschliche Geist ein System, in dem semantisch gehaltvolle Symbole auf gehaltkonforme Weise syntaktisch verarbeitet werden. Ein derartiges Symbolsystem kann aber in einem rein physischen System realisiert sein, gegeben dass dieses hinreichend komplex und kausal in die Welt eingebettet ist. (Auf die kausale Einbettung werde ich im §5 zurückkommen.) Eine Cartesische *res cogitans* oder entsprechende nicht-physische Eigenschaften sind dazu nicht nötig. Die Computertheorie des Geistes erlaubt so eine durchgehend *naturalistische* Erklärung kognitiver Fähigkeiten wie des Denkens. Mentale Zustände und Prozesse kommen deswegen in der Welt vor, weil es hinreichend komplexe natürliche physische Systeme gibt, die kausal in die Welt eingebettet sind und entsprechende Symbolsysteme realisieren.

2. Searles Gedankenexperiment des Chinesischen Zimmers

Searles Argumentation richtet sich gegen eine im Kontext der Computertheorie verfochtene Teilthese. Diese lässt sich anhand von Newells und Simons „Hypothese vom Physischen Symbolsystem (*physical symbol system hypothesis*)“ gut herausarbeiten. Diese zum Traditionsbestand der Künstlichen Intelligenz gehörende, wenn auch der Sache nach – wie wir im §5 sehen werden – problematische Hypothese besagt:

Ein physisches Symbolsystem (*physical symbol system*) verfügt über die notwendigen und hinreichenden Mittel für generelles intelligentes Handeln. (Newell/Simon 1976, 87)

Wenn wir das Haben mentaler Zustände als Vorbedingung für intelligentes Handeln unterstellen⁶ und mit Newell und Simon nur komplexe und spezifisch organisierte Symbolverarbeitung als für intelligentes Handeln hinrei-

⁶ Obwohl sie selbst nur intelligentes Verhalten von Systemen im Sinn haben, sind Newell und Simon oft genau so verstanden worden. Searle 1980, 437# ist nicht allein, wenn er Newell die These zuschreibt, „the essence of the mental is the operation of a physical symbol system“.

chend betrachten,⁷ dann stecken in der „Hypothese vom Physischen Symbolsystem“ die folgenden beiden Annahmen:

NOT Ein System *S* hat nur dann mentale Zustände, wenn *S* ein physisch realisiertes Symbolsystem ist.

HIN Wenn ein *S* ein – komplexes und spezifisch organisiertes – physisch realisiertes Symbolsystem ist, dann verfügt *S* allein kraft dieses Umstandes über mentale Zustände.

Der These HIN zufolge ist Symbolverarbeitung für sich genommen hinreichend für das Vorliegen mentaler Zustände. Sobald ein physisches System hinreichend komplexe Symbolverarbeitung betreibt, haben wir es mit einem denkenden System zu tun. Dies ist die These, die Searle als grundverkehrt zurückweist. Seiner Überzeugung nach reicht Symbolverarbeitung *unter keinen Umständen* für das Vorliegen mentaler Zustände hin, wie komplex die Verarbeitung auch sein mag. Dies soll sein Gedankengang zeigen.

Stellen wir uns, so Searle, einen auf geschickte Weise programmierten Computer vor, der sich, wie es scheint, mit uns versiert auf Chinesisch unterhält. Nehmen wir weiter an, der Computer sei als Gesprächspartner von einem chinesischen Muttersprachler nicht zu unterscheiden. (Das Gespräch findet natürlich per Bildschirm und Tastatur statt.) Der so programmierte Computer produziert seine kompetent anmutenden Antworten auf unsere eingegebenen Fragen dabei auf die für Computer übliche Weise: Er analysiert die Eingaben syntaktisch, ruft Subroutinen auf, greift auf Datenbanken zu etc. – kurz, er betreibt Symbolverarbeitung im erläuterten Sinn. Ein solches System realisiert offenkundig ein komplexes Symbolsystem. Hätte ein solches System mentale Zustände? Genauer fragt Searle, *verstünde* der so programmierte Computer Chinesisch in dem Sinne, in dem Chinesen Chinesisch verstehen? Seine Antwort hierauf entwickelt Searle mithilfe des folgenden Gedankenexperiments:

[S]tellen Sie sich vor, Sie wären in ein Zimmer eingesperrt, in dem mehrere Körbe mit chinesischen Symbolen stehen. Und stellen Sie sich vor, dass Sie (wie ich) keine Wort Chinesisch verstehen, dass Ihnen allerdings ein auf Deutsch abgefasstes Regelwerk für die Handhabung dieser Chinesischen Symbole gegeben worden wäre. Die Regeln geben rein formal – nur mit Rücksicht auf die Syntax und nicht auf die Semantik der Symbole – an, was mit den Symbolen gemacht werden soll. Eine solche Regel mag lauten: „Nimm ein

⁷ “By ‘sufficient’ we mean that any physical symbol system of sufficient size can be organized further to exhibit general intelligence“ (Newell/Simon 1976, 87)

Kritzeln-Kratzen-Zeichen aus Korb 1 und lege es neben ein Schnörkel-Schnarkel-Zeichen aus Korb 2“. Nehmen wir nun an, dass irgendwelche anderen chinesischen Symbole in das Zimmer gereicht werden, und dass Ihnen noch zusätzliche Regeln dafür gegeben werden, welche chinesischen Symbole jeweils aus dem Zimmer herauszureichen sind. Die hereingereichten Symbole werden von den Leuten draußen „Fragen“ genannt, und die Symbole, die Sie dann aus dem Zimmer herausreichen, „Antworten“ – aber dies geschieht ohne Ihr Wissen. Nehmen wir außerdem an, dass die Programme so trefflich und Ihre Ausführung so brav sind, dass sich Ihre Antworten schon bald nicht mehr von denen eines chinesischen Muttersprachlers unterscheiden lassen. Da sind Sie nun also in Ihrem Zimmer eingesperrt und stellen Ihre chinesischen Symbole zusammen; Ihnen werden chinesische Symbole hereingereicht und daraufhin reichen Sie chinesische Symbole heraus. In so einer Lage, wie ich sie gerade beschrieben habe, könnten Sie einfach dadurch, was Sie mit den formalen Symbolen anstellen, kein bisschen Chinesisch lernen. (Searle 1986, 31)

Das Chinesische Zimmer mit der Person darin ist ein physisches System, das ein Symbolsystem realisiert. Die zeichenmanipulierende Person im Zimmer – nennen wir sie den *Homunkulus* – überführt Zeichenketten in Zeichenketten und produziert Ausgaben zu Eingaben nach Regeln, die allein die syntaktischen Eigenschaften der involvierten Zeichen betreffen. Die betreffenden Zeichen haben dabei eine Bedeutung, schließlich handelt es sich um Zeichen des Chinesischen. Versteht der Homunkulus nun Chinesisch? Kennt er die Bedeutung der von ihm manipulierten Zeichen? Searles Urteil ist eindeutig: Der Homunkulus versteht kein Wort Chinesisch. Soweit ich sehe, sind nahezu alle Denker mit Searle hier einig. Ganz gleich wie viele Zeichen der Homunkulus bearbeitet und wie ausgefeilt die von ihm beachteten Regeln sind, er gewinnt kein Verständnis der Bedeutungen der von ihm rein syntaktisch bearbeiteten Symbole.

Searle betrachtet das erreichte Resultat nur als Zwischenschritt. Seiner Ansicht nach ergeben sich daraus Konsequenzen, und erst diese machen den Pfiff seines Gedankengangs aus:

Der Witz der Geschichte ist nun schlicht folgender: weil Sie ein formales Computerprogramm ausführen, verhalten Sie sich aus der Sicht eines Außenstehenden so, als verstünden Sie Chinesisch – und dennoch verstehen Sie nicht ein Wort Chinesisch. Wenn aber die Ausführung eines passenden Computerprogramms *in Ihrem Fall* nicht ausreicht, um Chinesisch zu verstehen, dann reicht das auch bei *keinem anderen digitalen* Computer aus. (...) [D]enn kraft seiner Ausführung eines Programms hat kein digitaler Computer irgendetwas, das Sie nicht haben. Der Computer hat – genau wie Sie – nichts außer einem Programm für die Handhabung uninterpretierter chinesischer Symbole. (Searle 1986, 31f)

Searle folgert hier lediglich, dass unser dem äußeren Anschein nach des Chinesischen mächtige Computer die Chinesische Sprache nicht versteht, da er die Bedeutungen der manipulierten Symbole nicht erfasst. Im Auge hat er dabei jedoch eine ganz generelle These. Er fasst diese so:

Kein Computerprogramm kann jemals ein Geist sein, und zwar aus dem einfachen Grunde, weil ein Computerprogramm bloß syntaktisch ist und der Geist mehr als syntaktisch ist. Der Geist ist semantisch – semantisch in dem Sinne, dass er mehr hat als eine formale Struktur: er hat einen Gehalt. (Searle 1986, 30)

Laut Searle ist die oben formulierte These HIN also falsch. Wahr ist vielmehr das glatte Gegenteil: Symbolverarbeitung allein kann für das Vorliegen mentaler Zustände nicht hinreichen. Damit wäre der „Hypothese vom Physischen Symbolsystem“ die Grundlage entzogen. Da Searle diese Hypothese als Kern der Computertheorie des Geistes ansieht, folgert er, diese Theorie sei widerlegt.

Wie deutlich geworden sein sollte, ist das Gedankenexperiment des Chinesischen Zimmers nur ein Schritt in Searles versuchter Widerlegung der Computertheorie des Geistes. Searle selbst macht dies deutlich, wenn er seinen Gedankengang in einem Argument zusammenfasst, dessen Prämissen so lauten (vgl. Searle 1986, 37f.):

1. Hirn verursacht Geist.
2. Syntax reicht nicht für Semantik aus.
3. Computerprogramme sind vollständig durch ihre formale (oder syntaktische) Struktur definiert.
4. Ein Geist hat geistige Inhalte, und zwar semantische Inhalte.

Das Gedankenexperiment des Chinesischen Zimmers gibt lediglich die Rechtfertigung für die Prämisse 2 ab. Die anderen Annahmen hält Searle aus unabhängigen Gründen für überzeugend. Aus seinen Prämissen zieht Searle zunächst einmal die folgende Schlussfolgerung:

Kein Computerprogramm kann aus eigener Kraft einem System einen Geist geben. Ein Programm ist, kurz gesagt, kein Geist und reicht – für sich selbst genommen – nicht hin, um einen Geist zu haben. (Searle 1986, 38)

Dies ist die für unsere Betrachtung der Computertheorie des Geistes relevante Konsequenz. Sie ergibt sich bereits aus den Prämissen 2-4; Annahmen über die kausale Relevanz des Gehirns für den menschlichen Geist benötigen wir dazu nicht. Wenn wir zusätzlich die von Searle in 1 formulierte Idee

aufgreifen, unser Gehirn liege unseren mentalen Zuständen kausal zugrunde, können wir mit ihm eine zweite Schlussfolgerung ziehen:

Hirnfunktionen können Geist nicht ausschließlich dadurch verursachen, dass sie ein Computerprogramm realisieren. (Searle 1986, 39.)

Der Geist einer Person besteht demnach nicht schlicht in durch die physischen Strukturen ihres Gehirns realisierten Zuständen und Verarbeitungsprozessen. Laut Searle entwirft die Computertheorie des Geistes also auch ein grundfalsches Bild des Verhältnisses von Geist und Gehirn.⁸

4. *Gedankenexperimente in der Philosophie*

Searles Art und Weise der Argumentation ist keine Ausnahme. Gedankenexperimente finden sich in allen Bereichen der gegenwärtigen Philosophie. Man denke nur an die Gettier-Fälle (vgl. Gettier 1963), Jacksons Geschichte von Mary (vgl. Jackson 1982) oder Putnams Reisen zu *Twin Earth* (vgl. Putnam 1975). Diese Gedankenexperimente teilen zentrale Eigenschaften.⁹ Erstens enthalten sie einen *narrativen Kern* – eine Geschichte, in der eine bestimmte, als möglich präsentierte Situation geschildert wird. Typischerweise betreffen die Vorgänge in der Situation dabei eine oder mehrere Personen. So beschreibt beispielsweise Jackson eine Situation, in der eine Person namens ‚Mary‘ in einer rein schwarz-weißen Umgebung aufwächst, nie etwas Farbigen sieht und trotzdem vollständiges physikalisches Wissen über die Welt erwirbt. Zweitens wird von uns ein *Urteil* über einen bestimmten Sachverhalt in der als möglich präsentierten Situation erwartet. So fragt uns Jackson: Nehmen wir an, Mary verlässt ihre schwarz-weiße Umgebung und sieht eine rote Rose. Lernt sie dann etwas? Das von uns eingeforderte Urteil ist zumeist als ein *modales Urteil* – dazu gleich mehr – zu verstehen, das wir auf Grundlage unserer Reflexion über die im narrativen Kern beschriebene mögliche Situation bilden sollen. Dabei wird drittens angenommen, dass die beschriebene Situation unser Urteil *begründet*. Gedankenexperimente dienen nicht einfach der Veranschaulichung.¹⁰ Unsere Reflexion über mögliche Situationen wird vielmehr als ein epistemisches Verfahren verstanden, das geeignet ist, uns zu wohl begründeten Urteilen zu führen und damit philosophische Positionen zu stützen oder zu widerlegen. Viertens ist ein Gedan-

⁸ Wie ist Searle zufolge das Verhältnis richtig bestimmt? Vgl. dazu Searle 1992.

⁹ Für eine ausführlichere Darstellung dieser Analyse vgl. Nimitz 2010.

¹⁰ Der Veranschaulichung dienende Geschichten wie Platons Höhlengleichnis sind folglich keine Gedankenexperimente.

kenexperiment typischerweise in einen *größeren Argumentationsgang* eingebettet, in dem aus dem durch das Gedankenexperiment begründeten Urteil Konsequenzen für philosophische Positionen gezogen werden. So folgert Jackson aus dem Umstand, dass die physikalisch allwissende Mary beim Anblick der roten Rose etwas lernt, der Physikalismus sei falsch.¹¹ Allerdings sollten wir die aus einem Gedankenexperiment gezogenen Konsequenzen nicht zum Gedankenexperiment selbst schlagen und klar zwischen dessen Resultat und den weiteren Konsequenzen unterscheiden.

Bereits diese Skizze sollte deutlich gemacht haben, dass wir es bei Gedankenexperimenten in der Philosophie nicht mit Experimenten im Wortsinne zu tun haben – also mit kontrollierten Manipulationen von Ereignissen, mit deren Hilfe zu Theorien oder Hypothesen bestätigende oder widerlegende Beobachtungen erzeugt werden sollen (vgl. Blackburn 1994, 131). Ein Gedankenexperiment in der Philosophie ist vielmehr die Bildung eines Urteils aufgrund der Reflexion über eine als möglich dargestellte Situation. Diese Urteilsbildung lässt sich schematisch als ein modaler Gedankengang in drei Schritten begreifen. Dies lässt sich am Fall des Chinesischen Zimmers gut deutlich machen. In einem ersten Schritt wird eine Situation als möglich dargestellt. Diese *Möglichkeitsprämisse* lautet in Searles Gedankenexperiment so:

- (1) Jemand könnte in der Situation des Homunkulus sein.

Die Geschichte vom Chinesischen Zimmer beschreibt, so wird unterstellt, eine mögliche, und keine unmögliche Situation. Wir stoßen also nicht auf Widersprüche oder sind gezwungen, notwendige Wahrheiten zu verletzen, wenn wir den narrativen Kern des Gedankenexperiments konsequent weiter ausformulieren und wohlwollend Details hinzufügen. Im zweiten Schritt wird eine modale Folgerung aus der Möglichkeitsprämisse gezogen. Dieses *modale Konditional* lautet in Searles Fall plausiblerweise so:

- (2) Wenn jemand in der Situation des Homunkulus wäre, dann würde er kein Wort Chinesisch verstehen.

Die genaue Art des modalen Konditionals variiert von Gedankenexperiment zu Gedankenexperiment. Dass wir es in Searles Fall mit einem kontrafaktischen Konditional zu tun haben, legt er selbst an folgender Stelle nahe:

¹¹ Zum Physikalismus vgl. Nimitz 2009.

In so einer Lage, wie ich sie gerade beschrieben habe, *könnten* Sie einfach dadurch, was Sie mit den formalen Symbolen anstellen, kein bisschen Chinesisch lernen. (Searle 1986, 31, meine Hervorhebung)

Aus der Möglichkeitsprämisse und dem modalen Konditional wird im dritten Schritt eine *modale Konklusion* gezogen. Bei Searle lautet diese:

- (3) Kein Homunkulus kann allein aufgrund seiner syntaktischen Manipulation Chinesischer Symbole deren Bedeutung verstehen.

Die Resultate philosophischer Gedankenexperimente sind zumeist Möglichkeitsbehauptungen. So folgert Gettier, jemand könne die gerechtfertigte wahre Meinung haben, dass *p*, ohne zu wissen, dass *p*, und Jackson schließt, es sei möglich, dass eine physikalisch allwissende Person etwas lernt. Searles Konklusion ist dagegen eine *Unmöglichkeitsbehauptung*. Allein aus Symbolmanipulation kann kein Verstehen der manipulierten Symbole resultieren – so Searle. Genau ein solches Resultat benötigt Searle für seine Argumentation. Genau betrachtet formuliert HIN nämlich die Möglichkeitsbehauptung, ein (komplexes etc.) Symbolsystem *würde* über mentale Zustände verfügen. Und diese These will Searle ja zurückweisen.

In einem Gedankenexperiment führt uns die Reflexion über eine als möglich dargestellte Situation vermittelt eines modalen Gedankengangs zu einem modalen Urteil. Dieser modale Charakter erklärt auf der einen Seite den Pfiff philosophischer Gedankenexperimente. Von Philosophen formulierte Theorien über z.B. Wissen, Bewusstsein, Bedeutung, Farben, Eigenschaften oder den freien Willen gehen typischerweise mit dem Anspruch einher, die Natur des betreffenden Phänomens anzugeben – also zu sagen, worin das Phänomen wesentlich besteht. Solche Theorien haben modale Konsequenzen. Wer beispielsweise behauptet, Wissen sei nichts anderes als gerechtfertigte wahre Meinung, legt sich auf die modale These fest, eine gerechtfertigte wahre Meinung sei notwendigerweise Wissen. Diese Notwendigkeitsbehauptung lässt sich durch den Nachweis der gegenteiligen Möglichkeit widerlegen. Genau darauf zielt Gettier mit seinen Gedankenexperimenten. Gedankenexperimente können also deswegen philosophische Theorien widerlegen oder stützen, weil diese zumeist modale Konsequenzen haben. Die Popularität von Gedankenexperimenten in der Philosophie kommt also nicht von ungefähr.

Auf der anderen Seite macht der modale Charakter des Gedankenexperimentierens deutlich, warum umstritten ist, ob wir es hier überhaupt mit einer akzeptablen Erkenntnismethode zu tun haben. Eine akzeptable Erkennt-

nismethode muss verlässlich zu wahren Überzeugungen führen. Ob das Gedankenexperimentieren eine akzeptable Erkenntnismethode ist, hängt folglich davon ab, ob wir durch Reflexion über als mögliche dargestellte Situationen verlässlich zu wahren Urteilen gelangen. Nun mag uns zwar der narrative Kern eines Gedankenexperiments gute Gründe für die Annahme geben, die beschriebene Situation sei eine mögliche. Aber was begründet den Dreh- und Angelpunkt des modalen Gedankengangs – das modale Konditional? Was begründet beispielsweise unser Urteil, jemand, der in der Situation des Homunkulus wäre, würde kein Wort Chinesisch verstehen? Kritiker des Gedankenexperimentierens sehen hier gar keine solide Grundlage, und die Anhänger dieser Methode sind uneins.¹² Populär sind die Ansichten, unser Wissen um die Wahrheit modaler Konditionale beruhe auf empirischem Hintergrundwissen zusammen mit alltäglichen Fähigkeiten (wie Williamson behauptet), sie beruhe auf *a priori*-Einsichten in Möglichkeit und Notwendigkeit (wie Lowe glaubt), oder ihr läge Wissen um begriffliche Beziehungen zugrunde (wie Jackson denkt).¹³ Für das von ihm angestrebte Gedankenexperiment scheint Searle sich auf die dritte dieser Optionen festzulegen. So klassifiziert er die durch das Gedankenexperiment gestützte These, Syntax reiche für Semantik nicht aus, als eine „begriffliche Wahrheit“ (Searle 1986, 38). Unser Urteil über die Situation des Homunkulus beruht demnach auf der begrifflichen Einsicht, dass etwas rein Formales wie die Manipulation von Zeichen kein Verstehen mit sich bringen kann, da Verstehen schon dem Begriff nach im Erfassen eines Inhalts besteht.

5. *Widerlegt Searle die Computertheorie des Geistes?*

Das Resultat des Gedankenexperiments des Chinesischen Zimmers lässt sich so auf den Punkt bringen: Die zeichenmanipulierende Person im Chinesischen Zimmer versteht kein Wort Chinesisch; überhaupt kann kein Homunkulus allein aufgrund seiner syntaktischen Manipulation Chinesischer Symbole deren Bedeutung verstehen. Dieses Urteil ist gut durch die angestrebte Überlegung gestützt und wird nahezu einhellig akzeptiert. Das Gedankenexperiment des Chinesischen Zimmers für sich betrachtet ist also erfolgreich – es zeigt, was es zu zeigen vorgibt. Umstritten ist, was daraus folgt.¹⁴ Laut Searle widerlegt sein Resultat die Computertheorie des Geistes ebenso wie jede funktionalistische Theorie, die mentale Zustände an kausa-

¹² Für Kritik vgl. Wilkes 1993, Kap. 1 und die Aufsätze in Knobe/Nichols 2008.

¹³ Vgl. Williamson 2007, Lowe 2002, Kap. 1 und Jackson 1998.

¹⁴ Vgl. Searle 1982, §5.1–5.5, die Beiträge in Preston/Bishop 2002 und Cole 2009.

len Rollen festmacht. Viele Philosophen widersprechen aufs Heftigste. Ihrer Einschätzung nach entwickelt Searle zwar ein erfolgreiches Gedankenexperiment. Aber dessen Resultat ist problemlos mit einer recht verstandenen Computertheorie des Geistes vereinbar; schon gar nicht untergräbt es den Funktionalismus. Der unter der Überschrift des Chinesischen Zimmers ausgetragene Streit betrifft also weder die Zulässigkeit noch das Resultat des Gedankenexperiments, sondern allein dessen Konsequenzen. Die Relevanz des Gedankenexperiments entscheidet sich folglich daran, ob Searle zu Recht aus seiner Einsicht über unseren Homunkulus auf die Unhaltbarkeit der Computertheorie des Geistes schließt.

Daran lassen sich auch dann erhebliche Zweifel anmelden, wenn wir semantische Fragen zunächst ignorieren. Der Homunkulus im Chinesischen Zimmer, so zeigt Searle überzeugend, versteht kein Wort Chinesisch. Warum soll daraus folgen, dass das System, zu dem der Homunkulus und sein Zimmer gehören, nicht *als Ganzes* entsprechende mentale Eigenschaften hat? Die Computertheorie begreift den menschlichen Geist als ein komplexes Symbolsystem. Dem gängigen Paradigma zufolge ist unser Geist in Verarbeitungseinheiten – sogenannte ‚Module‘ – gegliedert, die wie das ‚Wahrnehmungsmodul‘ oder das ‚Entscheidungsmodul‘ je spezifischen Funktionen dienen und ihrerseits wieder aus funktionalen Submodulen bestehen. Die Art der Organisation soll erklären, wie ein mechanisch symbolverarbeitendes System genuin intelligente Leistungen vollbringen kann: Seine intelligenten Leistungen beruhen darauf, dass die zu erbringenden Aufgaben in einfachere und wieder einfachere unterteilt sind, bis wir bei solchen anlangen, die auf simple mechanische Weise erledigt werden können. Die Leistungen sind dabei zumeist Leistungen des Systems, nicht seiner Teile. Das System nimmt wahr, nicht das Wahrnehmungsmodul.

Vor diesem Hintergrund erscheint Searles Schluss von Eigenschaften des Homunkulus auf die des Gesamtsystems ein, wie John Haugeland es nennt, „Teil-Ganzes-Fehlschluss“ (vgl. Haugeland 2002, 380) zu sein. Gegen diesen als „Systemeinwand“ bekannten Einwurf (vgl. Searle 1982, §5.1) verteidigt sich Searle, indem er uns vorstellen lässt, unser Homunkulus *internalisiere* das im Chinesischen Zimmer realisierte Symbolsystem. Auch in diesem Fall verstünde der Homunkulus laut Searle kein Wort Chinesisch; Gleiches gelte folglich für das von ihm internalisierte Symbolsystem. Diese Überlegung ist aber schon für sich genommen wenig überzeugend. Es ist nur schwer vorstellbar, unser Homunkulus *internalisiere* das ganze beschriebene modulare Symbolsystem, und nicht nur das Subsystem des Chi-

nesischen Zimmers. Dazu entwirft Searle hier ein vollständig neues und weitaus problematischeres Gedankenexperiment, dessen modales Konditional sich sicher nicht einfach aus begrifflichen Einsichten ergibt.

Gänzlich unplausibel wird Searles Antwort, wenn wir sie mit dem als „Robotereinwand“ bekannten Argument konfrontieren.¹⁵ Stellen wir uns einen autonom mobilen Roboter vor, der (i) per Kamera und Greifarm mit der Welt interagiert, der sich (ii) von außen betrachtet wie eine normale Person verhält – wie es scheint verändert er sein Tun, um erwarteten Nutzen zu maximieren, gibt anscheinend verständliche Antworten auf Fragen etc. – und von dem wir (iii) wissen, dass seine interne Verarbeitung derjenigen entspricht, die Psychologen bei menschlichen Denkern ausmachen – sie ist modular und seine Verarbeitungsmuster wie z.B. seine induktiven Schlüsse entsprechen denen, die Psychologen menschlichen Denkern zuschreiben. Würden wir einem solchen System nicht mit der gleichen Berechtigung mentale Zustände zuschreiben wie gewöhnlichen Menschen und z.B. ernsthaft sagen wollen, es habe den heranrauschenden Zug *gesehen* und *entschieden*, besser einen Meter zurückzurollen? Ich denke schon. Dazu sollte offenkundig sein, dass Searles Internalisierungsreplik in diesem Fall nicht verfängt.

Aber wenden wir uns der Frage der Semantik und damit dem Kern der Überlegung Searles zu. Searle argumentiert so: Mentale Zustände sind wesentlich Zustände mit einem semantischen Gehalt. Syntax allein reicht aber für semantischen Gehalt nicht aus, wie das Gedankenexperiment des Chinesischen Zimmers zeigt. Allein deswegen, weil ein *S* ein (komplexes etc.) physisch realisiertes Symbolsystem ist, kann *S* also keine mentalen Zustände haben. Dies klar zu begreifen war Anfang der 1980er Jahre hellsichtig. Bis zu diesem Zeitpunkt hatte die Zunft Fragen der Semantik oft links liegen gelassen. Es ist kein Zufall, dass Newells und Simons „Hypothese vom Physischen Symbolsystem“ die semantischen Eigenschaften der verarbeiteten Symbole überhaupt nicht berücksichtigt. Oder die Zunft hatte sich darauf beschränkt, unsere Deutung von außen als für deren Interpretation hinreichend anzusehen. Als Erklärung des Zustandekommens semantischer Gehalte greift das aber offenkundig zu kurz. Searle hat also Recht: Syntax allein reicht für Semantik tatsächlich nicht hin. Widerlegt das die Computertheorie des Geistes? Dem ist nicht so. Erinnern wir uns an die Grundidee dieser Theorie: Der menschliche Geist ist ein System, in dem semantisch

¹⁵ Vgl. Searle 1982, §5.2 sowie Rey 2002, 209–213 und Beckermann 1988, §5.

gehaltvolle Symbole auf gehaltkonforme Weise syntaktisch verarbeitet werden. Ein Verfechter dieser Idee muss keineswegs den semantischen Gehalt der im Geist verarbeiteten Symbole auf ihre syntaktischen Eigenschaften zurückführen. Genau das aber unterstellt Searle.

Nun argumentierte Searle nicht einfach an der Computertheorie des Geistes vorbei. Er hat vielmehr einen wichtigen Punkt im Blick. Die Anhänger der Computertheorie des Geistes wollen erklären, wie in unserer fundamental physischen Welt mentale Zustände vorkommen können. Die Computertheorie *allein* kann dies jedoch gar nicht leisten. Diese Theorie besagt, der menschliche Geist beruhe auf gehaltkonformer Symbolverarbeitung. Sie erklärt jedoch nicht, wie die semantischen Gehalte der verarbeiteten Symbole zuallererst zustande kommen. Ein Anhänger der Computertheorie des Geistes muss folglich zusätzlich zur Computertheorie, die eine *Theorie mentaler Zustände und -prozesse* ist, eine *Theorie des Inhalts* vorweisen können, die genau dies leistet. Searle ist dies klar. Als mögliche Theorie des Zustandekommens semantischen Gehalts zieht er jedoch allein die von ihm widerlegte Annahme in Betracht, semantischer Gehalt ergäbe sich aus syntaktischen Eigenschaften. Dies mag ganz im Sinne der „Hypothese vom Physischen Symbolsystem“ sein. Aber diese Hypothese ignoriert ja gerade das Problem der semantischen Eigenschaften. Außerdem hängen die einer Computertheorie verpflichteten Philosophen wie z.B. Fodor einer ganz anderen Grundidee an. Sie denken, die semantischen Eigenschaften der verarbeiteten Symbole ergäben sich aus kausalen Beziehungen zwischen Vorkommnissen von Symbolen im System und Umständen in der Welt.¹⁶

Diese auf verschiedene Weise ausbuchstabierte Grundidee lässt sich gut am Beispiel einer mechanischen Benzinuhr im Auto vor Augen führen. Die Zeigerstellungen einer solchen Benzinuhr tragen semantische Gehalte. Wenn beispielsweise der Zeiger auf „1/2“ deutet, zeigt dies an, dass der Tank halbvoll ist. Warum aber tragen die Zeigerstellungen semantische Gehalte? Weil durch eine kausale Verbindung gewährleistet ist, dass die Anzeige mit dem Füllstand des Tanks kovariiert und so die Zeigerstellung als *Indikator* für den Tankinhalt fungiert. Auf diese Weise kann eine rein kausale Verbindung das Vorliegen semantischer Eigenschaften erklären. Natürlich ist die Verknüpfung im Fall der Benzinuhr von uns künstlich so eingerichtet worden. Aber der Zeigerstand wäre auch dann ein Indikator der Füllmenge, wenn sich die kausale Verbindung durch einen natürlichen Pro-

¹⁶ Für eine Übersicht siehe Beckermann 2008, Kap. 12.

zess entwickelt hätte. Genau dies war, so die Anhänger der Computertheorie, bei natürlich evolvierten symbolverarbeitenden Systemen wie bei uns der Fall.

Auf der Grundlage seines Gedankenexperiments des Chinesischen Zimmers folgert Searle zu Recht, dass Syntax für Semantik nicht hinreicht. Damit erweist er die ins Visier genommene These HIN als falsch. Als Kritik an der „Hypothese vom Physischen Symbolsystem“ ist Searles Überlegung folglich überzeugend. Die Anhänger einer Computertheorie des Geistes kann sie allerdings nicht in Verlegenheit bringen; schon gar nicht widerlegt Searles Überlegung den Funktionalismus. Der Sache nach sind die Anhänger einer Computertheorie auf eine syntaktisch gelesene „Hypothese vom Physischen Symbolsystem“ gar nicht festgelegt. Sie hängen einer stärkeren Variante der These HIN an, die sich so auf den Punkt bringen lässt:

HIN⁺ Wenn *S* ein komplexes und spezifisch organisiertes physisch realisiertes Symbolsystem ist, *dessen Symbolvorkommnisse in systematischen kausalen Beziehungen zu Umständen in der Welt stehen*, dann verfügt *S* kraft dieser Umstände über mentale Zustände.

Diese These liefert die theoretische Begründung für die Idee, wir würden dem oben beschriebenen Roboter mit voller Berechtigung und ernsthaft mentale Zustände zuschreiben. Dazu nimmt HIN⁺ die funktionalistische Kernidee auf, mentale Zustände müssten über kausale Rollen erklärt werden, und ermöglicht die benötigte Erklärung des Zustandekommens semantischen Gehalts. Eine Widerlegung von HIN⁺ würde folglich Computertheorie und Funktionalismus gleichermaßen in arge Bedrängnis bringen. Aber diese These widerlegt Searles Überlegung des Chinesischen Zimmers nicht. Unter dem Strich bleibt Searle damit ein überzeugendes Argument schuldig. Nun darf man nicht verschweigen, dass Searle selbst dies anders sieht (vgl. Searle 1982, §5.2). Er hält den Verweis auf kausale Beziehungen für irrelevant und denkt, sein Gedankengang erweise HIN und HIN⁺ gleichermaßen als falsch. Immerhin sei es für den Homunkulus im Chinesischen Zimmer unerheblich, ob die vorkommenden Symbole kausal mit Umständen in der Welt kovariierten oder nicht; auch wenn sie dies täten, würde er aufgrund seiner syntaktischen Symbolverarbeitung kein Wort Chinesisch lernen. Das ist wieder richtig. Aber darin einen Einwand gegen die Computertheorie des Geistes zu sehen, hieße endgültig, einem Teil-Ganzes-Fehlschluss aufzusitzen.

III.

- Beckermann, A.: *Analytische Einführung in die Philosophie des Geistes*, Berlin 2008.
- Beckermann, A.: Sprachverstehende Maschinen. Überlegungen zu John Searles Thesen zur Künstlichen Intelligenz. In: *Erkenntnis* 28, 1988, 65–85.
- Blackburn, S.: *Oxford Dictionary of Philosophy*, Oxford 1994.
- Cole, D.: The Chinese Room Argument. In: *Stanford Encyclopedia* 2009.
- Fodor, J.: *LoT 2 – The Language of Thought Revisited*, Oxford 2008.
- Fodor, J.: *Psychosemantics*, Cambridge (Mass.) 1987.
- Gettier, E. L.: Is Justified True Belief Knowledge? In: *Analysis* 23, 1963, 151–153.
- Haugeland, J.: *Künstliche Intelligenz – Programmierte Vernunft?* Hamburg 1987. (Übers. von ders.: *Artificial Intelligence – the Very Idea*, Cambridge (Mass.) 1985)
- Haugeland, J.: Syntax, Semantics, Physics. In: Preston/Bishop 2002, 379–392.
- Horst, S.: The Computational Theory of Mind. In: *The Stanford Encyclopedia of Philosophy*, 2009.
- Jackson, F.: Epiphenomenal Qualia. In: *Philosophical Quarterly* 32, 1982, 127–136.
- Jackson, F.: *From Metaphysics to Ethics. A Defense of Conceptual Analysis*. Oxford 1998.
- Knobe, J. und Nichols, S. (Hgg.): *Experimental Philosophy*, Oxford 2008.
- Kurzweil, R.: *The Age of Intelligent Machines*, Cambridge (Mass.) 1990.
- Lowe, E.J.: *A Survey of Metaphysics*, Oxford 2002.
- Newell, A. und Simon, H.A.: Computer Science as Empirical Enquiry: Symbols and Search. In: *Mind Design II*, hrsg. von J. Haugeland, Cambridge (Mass.) 1997, 81–110.
- Nimtz, C.: ‚Physisches‘ und Multi-Realisierbarkeit, oder: zwei Probleme für den Physikalismus gelöst, in: *Physikalismus – Willensfreiheit – Künstliche Intelligenz*, hrsg. von Backmann, M. und Michel, J., Paderborn 2009, 23–42.
- Nimtz, C.: Thought Experiments as Exercises in Conceptual Analysis. In: *Grazer Philosophische Studien* 81, 2010, 191–216.
- Preston, J. und Bishop, M. (Hgg.): *Views into the Chinese Room. New Essays on Searle and Artificial Intelligence*, Oxford 2002.
- Preston, J.: Introduction. In: Preston/Bishop 2002, 1–50.

- Putnam, H.: The Meaning of 'Meaning'. In: Putnam, H.: *Mind, Language, and Reality. Philosophical Papers, Vol. 2*, Cambridge 1975, 215–271.
- Ravenscroft, I.: *Philosophie des Geistes. Eine Einführung*, Stuttgart 2008. (Übers. von ders.: *Philosophy of Mind. A Beginner's Guide*, Oxford 2005).
- Rey, G.: Searle's Misunderstanding of Functionalism and Strong AI. In: Preston/Bishop 2002, 201–225.
- Russell, S. und Norvig, P.: *Artificial Intelligence. A Modern Approach*, Upper Saddle River 2003.
- Searle, J.: Minds, Brains, and Programs. In: *Behavioral and Brain Sciences* 3, 1980, 417-457.
- Searle, J.: *Geist, Hirn und Wissenschaft*, Frankfurt 1986. (Übers. von ders.: *Minds, Brains, and Science*, Cambridge (Mass.) 1984)
- Searle, J.: Is the Brain a Digital Computer? In: *Proceedings and Addresses of the American Philosophical Association*, 64, 1990, 21–37.
- Searle, J.: *Die Wiederentdeckung des Geistes*, München 1993. (Übers. von ders.: *The Rediscovery of Mind*, Cambridge (Mass.) 1992.)
- Searle, J.: Twenty-One Years in the Chinese Room. In: Preston/Bishop 2002, 51–69.
- Wilkes, K.: *Real People. Personal Identity Without Thought Experiments*, Oxford 1993.
- Williamson, T.: *The Philosophy of Philosophy*, Oxford 2007.