

2 Klassische Kognitionswissenschaft

Für die Theorien der Kognitionswissenschaft, die hier als klassisch bezeichnet werden, sind zwei inhaltliche Punkte von zentraler Bedeutung: *Komputation* und *Repräsentation*. Auf die Komputation wird in den klassischen Erklärungen der Funktionsweise von Kognition verwiesen, während Repräsentationen zur Erklärung der Inhalte herangezogen werden. Hierbei geht also einmal um das *Wie* und einmal um das *Was* der kognitiven Verarbeitung. Der Geist soll, so die allgemeine Annahme, wie eine Art Computer funktionieren und dabei Informationen erhalten, verrechnen und ausgeben. Die verrechneten Informationen und Inhalte seien symbolische Repräsentationen, die zu einem bestimmten Moment im mentalen Leben eines Menschen auftauchen und zudem einen semantischen Gehalt besitzen. Mithilfe des Verweises auf die komputationale Verarbeitung der symbolischen Repräsentationen ließe sich, so die Überzeugung, schließlich erklären, wie der semantische Gehalt von Gedanken zustande kommt.

Sowohl die *komputationale Theorie* wie auch die *repräsentationale Theorie* des Geistes (im Folgenden: KTG und RTG) sind charakteristisch für ein Verständnis und die Erklärung von Kognition im Rahmen der klassischen Kognitionswissenschaft. Was es genauer mit den Begriffen von Komputation und Repräsentation hinsichtlich der Funktionsweise und der Inhalte des Geistes auf sich hat, und, wie beide Theorien zusammenhängen, soll im Folgenden genauer erläutert werden. Zudem soll herausgearbeitet werden, was die theoretischen Annahmen, konzeptionellen Mittel und explanatorischen Werkzeuge für Konsequenzen hinsichtlich eines Verständnisses einer kognitiven Architektur des Menschen haben. Dem klassischen Verständnis zufolge ist der Aufbau des menschlichen Geistes von einer Modularität geprägt, einer voneinander unabhängigen und funktional getrennten Verarbeitung entsprechend der verschiedenen kognitiven Fähigkeiten. Die folgende Darstellung richtet sich insbesondere auf die drei genannten Aspekte: die Funktionsweise der Verarbeitung, den Inhalt der Gedanken und den Aufbau des Geistes. Abschließend werden die historisch wichtigsten Gegenpositionen der klassischen Auffassung rekapituliert.

2.1 Komputationalismus und Repräsentationalismus

2.1.1 Die komputationale Theorie des Geistes

Seit den 1980er Jahren gilt die KTG als die populärste Theorie in der Kognitionswissenschaft und findet in allen Disziplinen vielfach Verwendung. Ganz allgemein formuliert liegt dieser Theorie die Idee zugrunde, dass der Geist in seiner

Funktionsweise einem digitalen Computer ähnelt: Es gibt einen Input, eine Verarbeitung und einen Output. Analog dazu erhalten wir als Menschen demnach sensorische Informationen, denken darüber nach und verhalten uns entsprechend. Währenddessen laufe die kognitive Verarbeitung wie eine Software auf einer physischen Hardware ab, wobei das Gehirn die Hardware dafür bietet. Durch die Entwicklung des Computers war bereits der konzeptionelle und technologische Hintergrund für die Formulierung einer derartigen Theorie des menschlichen Geistes vorgelegt. Zugleich konnten mit der komputationalen Theorie große Fortschritte bei der sich zu dieser Zeit schnell entwickelnden Forschung zur künstlichen Intelligenz gemacht werden. Wenn diese Theorie dazu dient, kognitive Fähigkeiten künstlich umzusetzen, schien ein Transfer zur Erklärung kognitiver Fähigkeiten bei biologischen Organismen naheliegend. So wurden schließlich die Erfolge in der KI zur Stützung der komputationalistischen Theorie angeführt und ihre Übertragbarkeit auf den menschlichen Geist angenommen.

Aber nicht nur die technologischen Entwicklungen waren wegweisend, sondern auch die zu dieser Zeit bereits entwickelten methodischen Mittel der Mathematik, der formalen Semantik und der symbolischen Logik (Frege 1884; Tarski 1935; Church 1936; Russell & Whitehead 1910, 1912, 1913 und viele weitere Arbeiten). Frege hatte eine Logik erster Ordnung entworfen und Tarski mit seiner Theorie der Referenz eine entsprechende Semantik geliefert. Mit der Entwicklung des Lambda-Kalküls konnte Church demonstrieren, dass es unentscheidbare Probleme gibt, da in einigen Fällen eine Antwort auf bestimmte Fragestellungen mathematisch nicht berechenbar ist. In Churchs Tradition befasste Turing (1936) sich weiter mit Überlegungen zum Halteproblem und der damit verbundenen Frage nach der Entscheidbarkeit, ob die Ausführung eines Algorithmus irgendwann zu einem Ende kommt oder unendlich abläuft. Sein Ergebnis war, dass es keinen Algorithmus gibt, mit dem diese Frage für alle möglichen Algorithmen und beliebigen Eingaben beantwortet werden kann. Er verwendete, um zu zeigen, dass das Halteproblem unentscheidbar ist, als Modell die nach ihm benannte Turing-Maschine. Die Turing-Maschine ist fähig, ein unendliches Speicherband, das in Felder unterteilt ist, durch einen Lese-Schreib-Kopf zu verändern. Das heißt, sie kann auf einem dieser Felder ein Symbol lesen, löschen, ein neues Symbol einfügen und dann zum nächsten Feld fortschreiten. Sie kann so effektiv alle Probleme berechnen, die ein Mensch mit Papier, Bleistift und Radiergummi lösen kann. Entsprechend dieser intuitiven Auffassung von Berechenbarkeit sind die nach einem bestimmten Algorithmus ablaufenden Funktionen zur Lösung bestimmter Aufgaben von der Turing-Maschine berechenbar, d.h. heute, sie sind ‚Turing-berechenbar‘. Damit ist dieser Berechenbarkeit zufolge alles, was prinzipiell und abgesehen von Kapazitätseinschränkungen auch von einem Menschen effektiv berechenbar ist, auch Turing-berechenbar.

Die entsprechende These, dass eine mathematische Funktion genau dann intuitiv berechenbar ist, wenn sie Turing-berechenbar ist, wurde später als Church-Turing-These bekannt (vgl. allgemein auch Turing 1948, 1950). Demnach können, wie geschildert, die Prozesse, die bei einer kognitiven Fähigkeit des Menschen, wie z.B. dem Rechnen, ablaufen durch die Turing-Maschine nachgeahmt werden. Entscheidend ist für uns an dieser Stelle die Aussage, dass alles, was der Mensch berechnen kann, auch von einer Maschine zu bewältigen ist. Dies eröffnete weiter die Sichtweise, dass der menschliche Geist funktional wie eine Maschine verstanden werden kann und die Implementation, wie z.B. in Form des Gehirns oder eben eines künstlichen physischen Werkzeugs, derweil nicht relevant ist (vgl. zur Formulierung der Church-Turing-These Kleene 1967: 232; vgl. für historische Überblicksdarstellungen auch Horst 1996: 28-34, 2011; Thagard 2005: 24).

Für Fodor ist das Modell der Turing-Maschine das Paradebeispiel für einen Computer, der verglichen mit dem menschlichen Geist analoge Fähigkeiten besitzt:

When philosophers of mind think about computers, it is often Turing machines that they are thinking about. And this is understandable. If there is an interesting analogy between minds qua minds and computers qua computers, it ought to be possible to couch it as an analogy between minds and Turing machines, since a Turing machine is, in a certain sense, as general as any kind of computer can be. (Fodor 1983: 38-39)

Die Entwicklungen in der Forschung zur KI machten später durch den Einfluss der KTG riesige Fortschritte bei den Versuchen, kognitive Kompetenzen wie Schlussfolgern, Sprechen, Problemlösen und Wahrnehmung maschinell zu implementieren.⁸ Während Newell, Shaw und Simon (1957, 1959) mit ihren frühen Arbeiten in diesem Bereich wichtige Grundlagen für komputationalistische Ansätze geliefert haben, wird als der philosophische Ideengeber der KTG häufig Putnam (1960, 1967) mit seiner Konzeption des (Maschinen-)Funktionalismus gesehen. Die philosophische Weiterentwicklung und Ausformulierung der KTG geht in der darauf folgenden Zeit maßgeblich auf Fodor (1975, 1980, 1987, 1993, 2000) zurück. Der von Putnam formulierte Funktionalismus zeichnet sich durch die Annahme aus, dass mentale Zustände und Ereignisse komputationale Zustände des Gehirns sind und als „computational parameters plus relations to biologically characterized inputs and outputs“ (Putnam 1988: 7) definiert werden können. Entscheidend sei eigentlich nicht die Implementierung in das Gehirn, in die Hardware – „[W]e could be made of Swiss cheese and it wouldn’t matter“ für die algorithmische oder kognitive Verarbeitung (Putnam 1975: 291). Was entscheidend sei,

⁸ Hierbei ging es aber nicht unbedingt immer darum, die Mechanismen, die bei Menschen auftreten, zu kopieren und zu replizieren.

ist die funktionale Organisation, d.h. die Weise, in der die mentalen Zustände kausal mit anderen Zuständen in Verbindung stehen, zu den Inputs und Outputs. Gegenstände wie Teekannen oder Käse hätten nicht aus dem Grund keine mentalen Zustände, weil sie nicht aus dem entsprechenden Material gemacht sind, sondern weil sie nicht die entsprechende funktionale Organisation aufweisen. Die These, dass beispielsweise auch Schweizer Käse mit der angemessenen funktionalen Organisation mentale Zustände aufweisen kann, ist als die These der multiplen Realisierbarkeit bekannt geworden; mentale Zustände seien als funktionale Zustände unabhängig von der Implementierung in einer bestimmten physischen Grundlage, also multipel realisierbar.

2.1.2 Die repräsentationale Theorie des Geistes

Die KTG basiert also auf der Annahme, dass der Geist einem digitalen Computer ähnelt und die kognitive Informationsverarbeitung entsprechend in Form einer Komputation abläuft. Diese Annahme steht aber nicht für sich allein, sondern wird von einer repräsentationalistischen Theorie, der RTG, begleitet. Es gäbe „no computation without representation“, wie Fodor (1975: 34) pointierend zusammenfasst. Repräsentationen seien die Elemente, die komputational verarbeitet werden, und zugleich die Voraussetzung dafür, dass überhaupt eine Komputation stattfinden kann.

Um an dieser Stelle zunächst eine allgemeine Interpretation des Begriffs von Repräsentation vorzulegen, wird der Ausdruck „Repräsentation“ in dieser Arbeit des Weiteren (und in auch vielen anderen Zusammenhängen) für die Kennzeichnung eines Zustandes verwendet, der für einen anderen Zustand einsteht. Eine viel zitierte Charakterisierung, häufig übernommene und in unserem Kontext hilfreiche Auffassung von Repräsentation ist die von Haugeland:

A sophisticated system (organism) designed (evolved) to maximize some end (e.g., survival) must in general adjust its behavior to specific features, structures, or configurations of its environment in ways that could not have been fully prearranged in its design. [...] But if the relevant features are not always present (detectable), then they can, at least in some cases, be represented; that is, something else can stand in for them, with the power to guide behavior in their stead. That which stands in for something else in this way is a *representation*; that which it stands for is its *content*; and its standing in for that content is *representing* it. (Haugeland 1991: 62; Hervorhebungen im Original)

Nach dem klassischen Verständnis sind die Repräsentationen zudem symbolisch strukturiert, was später noch genauer erläutert wird. Kognitive Prozesse werden währenddessen in Form der Komputation als rein formale Symbolmanipulation verstanden, d.h., die Manipulation der Repräsentationen erfolgt ausschließlich auf

Grundlage der formalen, aber nicht der semantischen Eigenschaften. Um die klassische Annahme einer formal bedingten Manipulation von Symbolen in einem wörtlichen Sinne zu betonen, dient im Rekurs wiederum die von Fodor postulierte Parallele zwischen den Fähigkeiten der Turing-Maschine und den Fähigkeiten des menschlichen Geistes:

More precisely: if, as many of us now suppose, minds are essentially symbol-manipulating devices, it ought to be useful to think of minds on Turing-machine model since Turing machines are (again “in a certain sense”) as general as any symbol-manipulating device can be. (Fodor 1983: 39)

Neben Fodor nehmen auch Pylyshyn (1980, 1984) und andere Vertreter der klassischen Auffassung an, dass der Geist Repräsentationen hinsichtlich ihrer syntaktischen Eigenschaften verarbeitet. Die symbolische Ebene sei eine natürliche funktionale Ebene für sich, folge bestimmten gesetzesähnlichen Regeln und sei in Putnams Sinne unabhängig von der spezifischen physischen Realisierung. Die funktionale Unabhängigkeit einer Repräsentation von der Implementierung wird währenddessen als der wesentliche Unterschied zwischen kognitiven und physischen Phänomenen gesehen. Die Annahme einer autonomen symbolischen Ebene sei währenddessen aber in Übereinstimmung mit den grundlegenden Prinzipien der Komputation, wie sie beispielsweise in den damals maßgebenden Konzeptionen von künstlicher Intelligenz zu finden waren. Diese besondere repräsentationalistische und komputationalistische Sicht, wie auch die Annahme von Symbolen als Elementen der Verarbeitung, ist eine Gemeinsamkeit, die durchweg in der klassischen Kognitionswissenschaft vertreten wird (vgl. Fodor 1975, 2000, 2003, 2008, Fodor & Pylyshyn 1988, Marr 1982, Newell & Simon 1976). Wenn es um das Verständnis kognitiver Prozesse geht, war Fodors Slogan, dass die RTG „the only game in town“ (Fodor 2008: 113) sei, damit für eine bestimmte Zeit nach der Kognitiven Wende durchaus treffend. Wie diese symbolischen Repräsentationen und ihre komputationale Verarbeitung genau zu verstehen sind, wird nun anhand der prominentesten Untersuchungen dargestellt.

2.2 Klassische Projekte

2.2.1 Newell & Simon über ein *Physical Symbol System*

In den 1950er Jahren entwickelten Newell, Shaw und Simon (1957, 1958, 1959) Programme, die eine allgemeine Strategie für das Lösen formalisierter Probleme liefern sollten. Diese Programme sollten aber nicht nur logische Probleme lösen können, sondern zudem als Modell der intelligenten Informationsverarbeitung dienen, das sich auf die menschliche Intelligenz übertragen lässt. Damit lieferten sie

nicht nur frühe Beiträge zur Entwicklung von künstlicher Intelligenz, sondern legten zugleich auch explizit eine Theorie des menschlichen Denkens vor (vgl. Newell & Simon 1961a: 2016). Diesen Anspruch verfolgten sie mit theoretischen und konzeptionellen Mitteln, die später die Grundlage für die KTG und RTG bildeten. Die Ergebnisse ihrer Arbeiten fassten sie als *Physical Symbol System*-Hypothese hinsichtlich intelligenten Verhaltens zusammen, die Newell mit folgenden Worten formuliert:

Physical Symbol System Hypothesis: The necessary and sufficient condition for a physical system to exhibit general intelligent action is that it be a physical symbol system. *Necessary* means that any physical system that exhibits general intelligence will be an instance of a physical symbol system. *Sufficient* means that any physical symbol system can be organized further to exhibit general intelligent action. *General intelligent action* means the same scope of intelligence seen in human action: that in real situations behaviour appropriate to the ends of the system and adaptive to the demands of the environment can occur, within some physical limits. The hypothesis takes as given the identity of symbol systems and universal systems, and asserts their connection to rationality, a concept which did not enter into their formulation. (Newell 1980: 170)

Diese durch die damaligen Befunde und Entwicklungen motivierte Hypothese fand, wenn auch in etwas anderen Formulierungen, schon bereits in früheren Arbeiten Erwähnung (vgl. z.B. Newell & Simon 1976: 116). Im Kern besteht sie aber immer aus einer zweiteiligen Behauptung mit weitreichenden Implikationen hinsichtlich der Bedingungen von Intelligenz im Allgemeinen und der Möglichkeit von künstlicher Intelligenz im Speziellen: Ein physisches System von Symbolen weist die (1) notwendigen und (2) hinreichenden Bedingungen für intelligentes Handeln auf. Diese zwei Behauptungen lassen sich im Einzelnen wie folgt näher explizieren:

(1) Das physische Symbolsystem *als notwendige Bedingung* für Intelligenz:

Wenn etwas kein physisches Symbolsystem ist, ist es nicht zu intelligenten Handlungen fähig. Damit ist die These als eine Bestimmung hinsichtlich jeder kognitiven Architektur zu verstehen. Wenn beispielsweise Menschen fähig sind, intelligent zu handeln, dann ist auch der menschliche Geist ein physisches Symbolsystem.

(2) Das physische Symbolsystem *als hinreichende Bedingung* für Intelligenz:

Wenn man es schafft, ein physisches Symbolsystem zu konstruieren, kann man prinzipiell eine künstliche Intelligenz entwickeln.

Wie aber haben wir uns die Verarbeitung innerhalb eines physischen Symbolsystems genauer vorzustellen? Das System soll nach Newell & Simon zunächst aus

einer Menge von Symbolen – oder auch: physischen Mustern – bestehen. Diese Symbole treten weiter als Komponenten einer anderen Klasse auf, der Klasse der Ausdrücke (oder Symbolstrukturen). Auf diese Weise der Klassenzugehörigkeit ist die Symbolstruktur aus einer Zahl von Symbolen zusammengesetzt, die zudem in einer physischen Weise miteinander in Beziehung stehen. Zu jeder Zeit beinhaltet das System so eine Reihe von Ausdrücken, die als Vorkommnisse von Symbolen aufgrund ihrer physischen Relationen eine bestimmte Struktur bilden. Zusätzlich finden in diesem System Operationen an den Ausdrücken statt, um andere Ausdrücke hervorzubringen. Schließlich lässt sich ein physisches Symbolsystem als eine Maschine verstehen, die über die Zeit hinweg eine sich entwickelnde Sammlung von Symbolstrukturen produziert (vgl. Newell & Simon 1976: 116).

Im Sinne der Übertragbarkeit der *Physical Symbol System*-Hypothese auf alle kognitive Architekturen und damit auch die des Menschen heißt das für Newell & Simon, dass kognitive Fähigkeiten als Prozesse der Symbolmanipulation verstanden werden sollen:

We can postulate that the processes going on inside the subject's skin – involving sensory organs, neural tissues, and muscular movements controlled by neural signals – are also symbol-manipulating processes; that is, patterns in various encodings can be detected, recorded, transmitted, stored, copied, and so on, by the mechanism of this system. (Newell & Simon 1961a: 2012)

Auch bei biologischen Systemen mit Sensoren, Nerven und Muskeln würden wie bei einem artifiziellen Symbolsystem Prozesse ablaufen, bei denen komplexe Symbole manipuliert werden. Diese Prozesse für das Generieren und Transformieren komplexer Symbolstrukturen können darüber hinaus auch selbst durch Symbole und Symbolstrukturen innerhalb des Systems repräsentiert werden. Diese Transformation und Generation von Symbolstrukturen entsprächen dabei weiter bestimmten Regeln, die von einem kognitiven System, egal ob Mensch oder Maschine, beim Problemlösen befolgt werden.

Newell & Simon (1961, 1972) beziehen sich insbesondere auf die Fähigkeit, Probleme lösen zu können, da sie hierin das primäre Indiz für Intelligenz sehen. Bei ihrem Versuch, Problemlösen als entscheidende kognitive Fähigkeit in Form des *General Problem Solver* zu modellieren, gehen sie von einer Zweck-Mittel-Analyse aus, bei der zunächst der Unterschied zwischen dem gegenwärtigen Zustand und einem Zielzustand errechnet wird. Weiter wird dann innerhalb eines Programms diejenige Transformation festgelegt, die diesen Unterschied verringert, und ermittelt, ob diese Transformation auf den gegenwärtigen Zustand angewandt werden kann. Die Analogie zum menschlichen Problemlösen lässt sich im Sinne von Newell & Simon folgendermaßen verdeutlichen: Ausgehend von einem Ziel oder Zweck, wie beispielsweise dem Sieg bei einem Schachspiel, machen die Spieler eine gewisse Anzahl von möglichen Zügen, um ein definiertes

Ziel, den Sieg, zu erreichen. Die kognitive Leistung der Problemlösung besteht hierbei darin, dass die Spieler darüber entscheiden, welche Züge gemacht werden müssen, um schließlich zu siegen. Aus allen Zuständen, die von einem Ausgangszustand aus erreicht werden können, und der Menge der möglichen Transformationen (oder Schachzüge) während der Problemlösung (oder dem Weg zum Sieg) setzt sich der sogenannte *problem space*⁹ zusammen, also die Menge der möglichen Zustände inklusive dem Zielzustand (vgl. Newell & Simon 1976: 121 ff.). Das Lösen eines Problems ist somit eine Frage der Identifizierung eines Lösungszustandes und der diesbezüglichen Suche nach einem Pfad durch den Problemraum. Die Identifizierung und die Suche der Lösung, das eigentliche Problemlösen, ließen sich, so der entscheidende Punkt von Newell & Simon hinsichtlich ihrer Entwürfe einer künstlichen Intelligenz, durch die Symbolmanipulation in einem physischen Symbolsystem realisieren und dementsprechend als solches verstehen, also als Kombination und Transformation von physischen Symbolen innerhalb von Systemstrukturen. Die Idee, dass physische Symbole einer Systemstruktur entsprechend regelmäßig manipuliert werden, ist wegweisend für Fodors spätere sprachähnliche Konzeption von Gedanken und viele weitere Untersuchungen im Zeichen der KTG und der RTG, wie wir sie beispielsweise bei Marrs Erklärung der visuellen Wahrnehmung finden.

2.2.2 Marr über das Sehen

Marr (1982; vgl. auch Marr & Poggio 1976, 1977, 1979) unterscheidet bei der Beschreibung der visuellen Informationsverarbeitung zwischen drei Ebenen: einer komputationalen Ebene, einer algorithmischen Ebene und einer Ebene der Implementierung. Diese Unterscheidung ist für ihn dadurch motiviert, dass Kognition eine besondere Art der Erklärung verlangt. Während man Farbsehen durch physische Mechanismen erklären könne, müsste, so Marr, bei einer umfassenden Beschreibung von visueller Perzeption, die beispielsweise auch Objekterkennung beinhaltet, zwischen mehreren Ebenen unterschieden werden, um alle Phänomene adäquat zu erfassen:

[S]ome [problems] are related mainly to the physical mechanisms of vision—such as afterimages [...] or such as the fact that any color can be matched by a suitable mixture of three primaries [...]

⁹ Bermúdez benutzt hier den Ausdruck „search space“ (vgl. Bermúdez 2010: 150 ff.).

On the other hand, the ambiguity of the Necker cube seems to demand a different kind of explanation. (Marr 1982: 25)

Bei einem Kippbild wie dem Necker-Würfel kann es nämlich nicht der rein sensorische Input in Form der Retinareizung sein, der dafür verantwortlich ist, dass man den Würfel auf zwei verschiedene Weisen, nach vorne oder nach hinten geneigt, sehen kann, da es gerade der Input ist, der sich nicht verändert. Vielmehr sei dieser Effekt des Kippbildes auf den Output der Informationsverarbeitung zurückzuführen, die sich nach Marr nur mittels dreier Ebenen erschöpfend beschreiben lässt. Die unterschiedlichen Ebenen, die bei der Beschreibung der visuellen Informationsverarbeitung berücksichtigt werden müssen, lassen sich wie folgt charakterisieren (vgl. dazu Marr 1982: 24-27):

Komputationale Ebene: Auf der ersten (und höchsten) Ebene wird die komputationale Verarbeitung der Inputs und Outputs beschrieben, wobei der Output des menschlichen Sehens in einem gewissen Sinne reichhaltiger als der reine Input ist. Reicher ist der Output in dem Sinne, dass es die Aufgabe des visuellen Systems ist, auf der Input-Basis einer Retinareizung eine Repräsentation der 3-dimensionalen Form eines Objektes und seiner Position im Raum als Output zu erhalten, sodass man schließlich ein Objekt erkennen kann. Zudem hängt das Erkennen eines Objekts weiter mit der Fähigkeit zusammen, von einer bestimmten Perspektive auf ein perspektivisch unabhängiges Objekt zu extrapolieren, denn die Repräsentation der Form und des Ortes eines Objektes wird im Output als objektzentriert aufgefasst und verbleibt nicht allein im egozentrischen Referenzrahmen des Betrachters. Wenn sich beispielsweise zwei Personen gegenüberstehen und die eine Person die andere auffordert, ihr den Gegenstand in der rechten Hand zu reichen, hat sie dabei, wenn die aus ihrer Perspektive gesehen linke Hand des Gegenübers gemeint ist, einen von ihr unabhängigen Bezugsrahmen hinsichtlich der räumlichen Perspektive verwendet.

Ebene der Repräsentation und des Algorithmus: Auf dieser Ebene geht es um die Art und Weise der Informationsverarbeitung, um die Algorithmen, die das System von den Inputs zu den angemessenen Outputs führt. Die Ausgangspunkte der visuellen Informationsverarbeitung sind für Marr die Diskontinuitäten der Lichtintensität, da diese Auskunft über die Grenzen der Objekte und andere relevante physische Eigenschaften geben würden. Die Informationen über die Veränderungen der Lichtintensität würden weiter die Grundbestandteile der Repräsentationen bereitstellen. Auf der algorithmischen Ebene würden die verschiedenen Informationen in Form von Repräsentationen vereint, wobei die Verarbeitung, angefangen bei der Retinareizung bis zur räumlichen Repräsentation der Objekte, schrittweise abläuft: Wenn beim Sehen Informationen empfangen werden, geschehe dies auf der Basis von immer komplexer werdenden Repräsentationen ausgehend von dem

retinalen Bild, über einen 2-dimensionalen, ersten Entwurf (*primal sketch*), einen $2\frac{1}{2}$ -dimensionalen Entwurf (*$2\frac{1}{2}$ -dimensional sketch*) hin zu der schließlich 3-dimensionalen Repräsentation (*3-dimensional sketch*).

Auf der Basis des Bildes auf der Retina, das z.B. die Lichtintensität und Bereiche von relativer Helligkeit oder Dunkelheit repräsentiert, liefert der erste Entwurf Informationen über die grundlegende Geometrie des Sichtfeldes (Kantensegmente, Grenzen, Gruppen etc.). Das Resultat der weiteren Informationsverarbeitung ist der $2\frac{1}{2}$ -dimensionale Entwurf, bei dem bestimmte Informationen des ersten Entwurfes über jeden Punkt des Sichtfeldes hinsichtlich seiner Distanz zum Beobachter repräsentiert werden; der $2\frac{1}{2}$ -dimensionale Entwurf ist hierbei jedoch noch perspektivisch beobachterzentriert. Die stabile und perspektivenunabhängige Repräsentation der Objektform liefert erst der 3-dimensionale Entwurf (vgl. Marr 1982: 37).

Ebene der Implementierung: Bei dieser Ebene geht es um die Beschreibung dessen, auf welche Weise Repräsentationen und Algorithmus physisch realisiert sind. Diese Beschreibungsebene bezieht sich in diesem Fall auf die Physiologie des visuellen Systems.

Kurz gesagt: Dem visuellen Systems dienen nach Marr repräsentationale Grundelemente als Input, die in einer Reihe von komputationalen Schritten entsprechend einer algorithmischen Verarbeitung schließlich zu einem Output transformiert werden, der die 3-dimensionale Umgebung beobachterunabhängig repräsentiert. Marr räumt der komputationalen Ebene gegenüber der Implementierung und dem Algorithmus eine bestimmte Priorität ein, da erst auf dieser Ebene eine Untersuchung der Repräsentationen und des dazugehörigen Algorithmus möglich sei. Eine ähnliche Herangehensweise und Dreiteilung von Verarbeitungsstufen ist auch bei Fodors Language of Thought-Hypothese wiederzufinden, um speziell den semantischen Gehalt von Gedanken durch repräsentationale Elemente und deren regelgeleitete Verarbeitung auf neuronaler Basis zu erklären.

2.2.3 Fodors Language of Thought-Hypothese

Die KTG und die RTG wurden auch herangezogen, um die semantischen Eigenschaften von mentalen Zuständen zu erklären. Für diese Aufgabe formulierte Fodor die *Language of Thought*-Hypothese (LoTH), nach der Denken als komputationale Symbolmanipulationen in einem internen und sprachähnlichen Repräsen-

tationsmedium abläuft. Die „Language of Thought“ (LoT), die Sprache des Denkens/der Gedanken¹⁰, oder auch „Mentalesisch“ genannt, ist dabei aber nicht wie eine natürliche gesprochene Sprache zu verstehen, sondern als Repräsentationsformat von Begriffen. Der Erwerb einer natürlichen Sprache würde, so Fodor, bereits eine LoT voraussetzen, die selbst nicht erlernt wurde, sondern angeboren ist (vgl. Fodor 1975: 63 ff.). Beim Spracherwerb sei es so, dass wir eine Sprache erlernen, indem wir Hypothesen bilden und überprüfen, und dabei die semantischen Eigenschaften von Prädikaten und auch die Bedingungen lernen, unter denen eine bestimmte Prädikation zulässig ist. Die zugrundeliegende LoT sei währenddessen *„a metalanguage in which representations of object language predicates are formulated“* (Fodor 1975: 65; Hervorhebung im Original). Die offenen Sprachäußerungen übernehmen die semantischen Eigenschaften der Gedanken eines Sprechers, während die Gedanken wiederum die semantischen Eigenschaften von mentalen Repräsentationen übernehmen, die in der LoT ‚formuliert‘ bzw., genauer gesagt, formatiert sind (vgl. Fodor 1981: 26). Die semantischen Eigenschaften von Gedanken sollen dabei durch die repräsentationalen Elemente samt ihrem sprachähnlichen Format erklärt werden können.

Bevor durch die Protagonisten der Kognitiven Wende das Denken über den menschlichen Geist revolutioniert wurde, war der Behaviorismus die populärste Sicht in Psychologie und Philosophie. Im Behaviorismus geht man vom beobachtbaren Verhalten der Menschen (und nicht-menschlichen Tiere) aus, um dann auf die Zustände mentaler Natur oder auch introspektive Fähigkeiten zu schließen, da sich nur das Verhalten mit rein empirischen Methoden adäquat fassen lasse, aber nicht unmittelbar die geistigen Vorgänge selbst. Wissenschaftlich sollte der Geist aufgrund der fehlenden empirischen und damit intersubjektiven Überprüfbarkeit somit nicht der Ausgangspunkt einer psychologischen Untersuchung sein. Die LoTH ist das historisch schließlich erfolgreichere Gegenprogramm, bei dem es explizit um die Annahme interner mentaler Zustände und deren Bedeutungsgehalt geht. Fodors Projekt besteht darin, sowohl der alltagspsychologischen wie auch der naturwissenschaftlichen Sicht auf mentale Zustände gerecht zu werden. Vor einem naturalistischen Hintergrund soll dabei ein reduktiver Physikalismus, also

¹⁰ Beckermann (1997) weist daraufhin, dass der Ausdrücke „lingua mentis“ oder „language of thought“ zum ersten Mal wahrscheinlich von Harman (1973) verwendet wurde. Da sich die Ausdrücke der „Sprache des Geistes“, wie Beckermann sie nennt, nach Fodor auf interne physische Systemzustände beziehen, also z. B. auf bestimmte neuronale Feuermuster oder auch Bitmuster im Speicher eines Computers, empfiehlt er die Bezeichnungen „Sprache des Gehirns“ oder „Sprache des Computers“. Gegenüber Beckermanns Verwendung von „Sprache des Geistes“ wird hier „language of thought“ wörtlich als „Sprache des Denkens/der Gedanken“ übersetzt.

die These, dass alles auf physische Zustände zurückgeführt werden kann, in Form des Funktionalismus mit einer alltagspsychologischen Deutung der Intentionalität vereinbart werden (vgl. Fodor 1975: 1–27; Fodors gesamtes Projekt wird ausführlich in *Psychosemantics* (1987) dargelegt).

Gemäß der Alltagspsychologie geht Fodor von mentalen Zuständen aus, die als propositionale Einstellungen, wie z.B. als Überzeugungen, Wünsche und Urteile, im mentalen Leben einer Person auftauchen und einen semantischen Gehalt besitzen. Wenn man einer Person eine propositionale Einstellung zuschreibt, dann beschreibt man nach Fodor aber auch zugleich gehaltvolle Gehirnvorgänge dieser Person. Mit der Sichtweise, dass neuronale Zustände semantisch bewertbar sind, nimmt Fodor eine zwischen Naturalismus und Alltagspsychologie vermittelnde Position ein, die als sogenannter intentionaler Realismus bekannt ist (vgl. Fodor 1987: 151). Die semantische Bewertung soll der alltagspsychologischen Zuschreibung von mentalen Zuständen Rechnung tragen, während zugleich die psychologischen oder empirisch-wissenschaftlichen Theorien berücksichtigt werden, mit denen die kognitive Prozesse als „causal sequences of tokenings of representations“ (Fodor 1987: 16–17) beschrieben werden. Die LoTH bildet dabei den erläuternden und explanatorischen Hintergrund für die Auffassung des intentionalen Realismus.

Für eine Rekonstruktion der LoTH schlägt Schneider (2009: 281 f. und 2011: 11 ff.) folgende Formulierungen von Grundannahmen (1–3) vor, mit deren Hilfe sich später die Argumentation für die Existenz einer LoT leicht nachvollziehen lässt. Schneider setzt bei der gerade genannten naturwissenschaftlichen Beschreibung von Kognition an:

- (1) Kognitive Prozesse sind kausale Sequenzen von Vorkommnissen von Repräsentationen.

Menschliche Kognition wird als etwas verstanden, das in Form einer Manipulation von symbolischen Repräsentationen abläuft, die ihrerseits im Gehirn physisch realisiert sind und dort systematisch verarbeitet werden. Weiter soll eine Proposition nach Fodor die Struktur von sprachähnlichen, aus den symbolischen Repräsentationen zusammengesetzten Komplexen besitzen; diese bilden die Sätze der LoT. Durch die neuronale Realisierung im Gehirn einer Person *S* hätten die komplexen Repräsentationen eine syntaktische Struktur, während sie zudem einen semantischen Gehalt, die Proposition, dass *p*, aufweisen. Wenn *S* eine Überzeugung, dass *p*, hegt, sei das unmittelbare Objekt einer Überzeugung ein aus den einzelnen Repräsentationen zusammengesetztes, komplexes Symbol, und damit ein Satz in der LoT:

What's at issue, however, is the internal structure of these functionally individuated states. [...] only the *intentional objects* of mental states are complex. I think they constitute a language; roughly, the syntactic structure of mental states mirrors the semantic relations among their intentional objects. (Fodor 1987: 138)

Die repräsentationalen Zustände in Form von Symbolen konstituieren im Zusammenhang eines syntaktischen Regelsystems also eine Art von Sprache samt ihrer semantischen Dimension und können als Sätze der Form p ausgedrückt werden. Dass die Syntax derweil die Semantik „spiegeln“ würde, ist weiter erläuterungsbedürftig. Halten wir aber vorher fest, dass das, was geglaubt wird, die Proposition, zunächst ein Komplex aus Repräsentationen mit syntaktischen Eigenschaften und zugleich das intentionale Objekt der propositionalen Einstellung des Denkenden ist. Eine Überzeugung wird somit als eine dreistellige Relation zwischen dem Denkenden, dem mentalesischen Satz und der Proposition verstanden, und nicht nur zwischen dem Denkenden und der Proposition (vgl. auch Aydede 2010; Stich 1994).

Über den mentalesischen Satz wird von Fodor behauptet, dass er den semantischen Gehalt der Proposition im Sinne einer syntaktischen „Spiegelung“ aufweist. Anders ausgedrückt, der semantische Gehalt des propositionalen Ausdrucks soll durch die Vorkommnisse mentaler Repräsentationen wiedergegeben werden, die selbst nur eine syntaktische Struktur besitzen. Die Semantik würde, so Fodor, insbesondere durch die kombinatorischen Regeln aufgrund der syntaktischen Eigenschaften der Repräsentationen bestimmt:

(2) Die internen Repräsentationen besitzen eine (2.1) kombinatorische Syntax und (2.2) Semantik.

Die Repräsentationen bilden nach Fodor zunächst entsprechend ihrer Eigenschaften eine Syntax wie auch ein formales System. Damit wird eine Menge von kombinatorischen Ableitungsregeln bereitgestellt, die formal über die symbolischen Eigenschaften der Repräsentationen bestimmt sind. Diese kombinatorische Syntax ist damit strukturell und systematisch aus den komplexen (molekularen) Repräsentationen, den intentionalen Objekten, und einfachen (atomaren) Konstituenten aufgebaut (vgl. Fodor 1987: 135–137; Fodor & Pylyshyn 1988: 12 f.):

(2.1) Kombinatorische Syntax: Komplexe Repräsentationen sind aus atomaren Symbolen unter Berücksichtigung der syntaktischen Struktur aufgebaut.

Betrachten wir beispielsweise das Schlussfolgern als Fähigkeit, formal zu denken, um diese Annahme besser zu verdeutlichen: Beim Schlussfolgern werden nach Fodor Repräsentationen komputational verarbeitet. Hierbei sollen die intentionalen Zustände komplexe symbolische Repräsentationen beinhalten, die bestimmte

semantische Eigenschaften besitzen, die den formalen Regeln der Syntax gehorchen. Die Möglichkeit von Inferenzen, bei denen schließlich der semantische Wert entscheidend ist, sei dadurch gegeben, dass die Inferenzen empfänglich für die Syntax als formale Bedingung sind, aber nicht für semantische Intuitionen. Während einer Schlussfolgerung ergibt sich also die Semantik allein aus den strukturellen Bedingungen der Syntax:

What makes syntactic operations a species of formal operations is that being syntactic is a way of not being semantic. Formal operations are the ones that are specified without reference to such semantic properties as, for example, truth, reference and meaning. (Fodor 1981: 309)

Das heißt, die semantischen Eigenschaften von symbolischen Repräsentationen sind in syntaktischen Ableitungsregeln kodiert und Schlussfolgern ist ein Prozess, bei dem die syntaktischen Eigenschaften der Symbole als kausale Determinanten fungieren. Die syntaktischen Regeln bewahren dabei die semantischen Beziehungen, da mentale Operationen an internen Repräsentationen nur für die syntaktische Struktur der Symbole kausal empfänglich sind. Auf diese Weise werden bei der Symbolmanipulation auch die semantischen Eigenschaften der beinhalteten Gedanken erhalten. Mit Haugelands häufig zitierten Worten zusammengefasst: „If you take care of the syntax, then the semantics will take care of itself“ (Haugeland 1985: 106). So kommen wir zu der folgenden These über die Semantik innerhalb der LoT:

(2.2) Kombinatorische Semantik: Die Bedeutung oder der semantische Gehalt der Sätze in der LoT, der molekularen Repräsentationen, ist eine Funktion der Bedeutung seiner atomaren Symbole zusammen mit der syntaktischen Form.

Für Fodor ist die Frage nach der Erklärung der Semantik von Gedanken falsch ausgerichtet, wenn man dabei von der Semantik bzw. von bedeutungsgeladenen Intuitionen, aber nicht von den formalen Bedingungen ausgeht. Bei der LoTH wird jedoch von der Syntax bzw. der dadurch gegebenen Kombinatorik der repräsentationalen Elemente und zusammengesetzten Ausdrücke ausgegangen:

The question we're arguing about isn't, then, whether mental states have a semantics. Roughly, it's whether they have a syntax. Or, if you prefer, it's whether they have a *combinatorial* semantics: the kind of semantics in which there are (relatively) complex expressions whose content is determined, in some regular way, by the content of their (relatively) simple parts. (Fodor 1987: 138)

Die genannten Grundannahmen der LoTH lassen sich schließlich wie folgt zusammenfassen:

(3) Denken als eine Art der Symbolmanipulation bewahrt die semantischen Eigenschaften der beinhalteten Gedanken, da mentale Operationen auf internen Repräsentationen geleitet sind von der syntaktischen Struktur der Symbole.

Fodor mobilisiert eine Reihe von im Folgenden wiederzugebenden Argumenten, um seine Behauptung zu stützen, dass Gedanken sprachähnlich verfasst sind, d.h. also, dass mentale Repräsentationen in einem System mit Syntax und entsprechender Semantik auftreten. Genauer gesagt, würden sich die Merkmale von Gedanken wie Produktivität, Systematizität, Kompositionalität und Kohärenz auch nur dadurch erklären lassen, dass mit der LoT die Bedingungen hierfür geboten sind. Mit der kombinatorischen Syntax (und Semantik) soll man die Produktivität, Systematizität und Kompositionalität von Gedanken genauso erklären können, wie durch die Wörter und ihre Kombinationen innerhalb einer natürlichen Sprache. Das allgemeine und etwas verblüffende Argument im Hintergrund dafür, dass Gedanken eine sprachähnliche und kombinatorische Struktur besitzen, baut Fodor auf folgende Weise auf:

- (a) There's a certain property that linguistic capacities have in virtue of the fact that natural languages have a combinatorial semantics.
- (b) Thought has this property too.
- (c) So thought too must have a combinatorial semantics. (Fodor 1987: 148)

Das Problem von Fodors Argument ist ein offensichtlicher Fehlschluss – *non sequitur*: (a) weil natürliche Sprachen die Eigenschaft einer kombinatorischen Semantik haben, haben linguistische Fähigkeiten diese bestimmte Eigenschaft; (b) Gedanken haben diese Eigenschaft; also haben (c) Gedanken auch eine kombinatorische Syntax; oder kurz: $F(x)$, weil $G(x)$; gegeben $F(y)$; also $G(y)$. Zum einen folgt dies nicht und zum anderen liegt hier eine Äquivokation von ‚linguistischen Fähigkeiten‘ und ‚Gedanken‘ (im Allgemeinen) vor. Dies interessiert Fodor aber nicht, denn „one man's affirming the consequent is another man's inference to the best explanation“ (Fodor 1987: 149). Es reicht Fodor aus, dass er überzeugt ist, die kognitiven Phänomene und Fähigkeiten besser erklären zu können als andere Ansätze, sodass er nicht zeigen müsse, dass die LoTH auch notwendigerweise wahr sei. Die weiteren, nun folgenden Argumente (i), (ii) und (iii) sind ebenfalls in diesem rein pragmatischen Zusammenhang und in Anbetracht des behaupteten Fehlens einer besseren Erklärung zu verstehen. Dafür, dass allgemein eine LoT anzunehmen ist, würde schließlich der Erfolg der klassischen Theorien in der Kognitionswissenschaft wie eben der LoTH sprechen.

(i) *Die LoT als Bedingung für die Produktivität von Gedanken*

Fodor geht davon aus, dass man als Sprecher einer natürlichen Sprache fähig ist, auch Sätze zu verstehen, die man nie zuvor gehört hat. Prinzipiell können dies unendlich viele Sätze sein. Einen Satz zu verstehen, heißt zugleich, dass man den ausgedrückten Gedanken bzw. die Proposition erfasst hat, die der Satz ausdrückt. Wenn man unendlich viele Sätze verstehen kann, so lässt sich schließen, dass es auch unendlich viele Gedanken gibt, die man potentiell erfassen kann. Zwar ist es aufgrund unserer natürlichen Einschränkungen ausgeschlossen, dass wir diese Gedanken einmal haben werden oder gar als Vorkommnisse mentaler Repräsentationen speichern und abrufen können, jedoch ist dies nach Fodor wenigstens prinzipiell möglich. In genau dem Sinne, dass wir eine unbegrenzte Anzahl von Gedanken entwickeln können, werden Gedanken als produktiv verstanden (vgl. Fodor 1975: 31, 1987: 147 ff., 1998: 94 ff.; Fodor & Pylyshyn 1988: 116; siehe auch Aizawa 2003: 44 f.; Aydede 2010; Schneider 2011: 9 ff.).

Die Möglichkeit der Produktivität von Gedanken, auf der Basis eingeschränkter Bedingungen eine unbegrenzte Anzahl dieser entwickeln zu können, erklären Fodor & Pylyshyn durch Rekursion innerhalb eines Systems von Repräsentationen. Auch aus einer eingeschränkten Menge von Ausdrücken könnten durch unterschiedliche Zusammensetzungen verschiedenste und unendlich viele Propositionen aufgebaut sein:

[T]here are indefinitely many propositions which the system can encode. However, this unbounded expressive power must presumably be achieved by finite means. The way to do this is to treat the system of representations as consisting of expressions belonging to a generated set. More precisely, the correspondence between a representation and the proposition it expresses is, in arbitrarily many cases, built up recursively out of correspondences between parts of the expression and parts of the proposition. But, of course, this strategy can only operate when an unbounded number of the expressions are non-atomic. So linguistic (and mental) representations must constitute [symbol systems possessing combinatorial syntax and compositional semantics]. (Fodor & Pylyshyn 1988: 33; Ergänzung im Sinne ihrer Fussnote 8)

Auch wenn Menschen nicht über unendlich viele der Repräsentationen verfügen, reiche es aus, wenn sie über ein System verfügen, das die Konstruktion von unendlich vielen Gedanken auf der Basis von begrenzt vielen atomaren Elementen ermöglicht. Das einzige System, dass die Produktivität von Gedanken garantiert, soll ein repräsentationales Symbolsystem wie die LoT mitsamt einer kombinatorischen Syntax und Semantik sein. Auch im Folgenden laufen die Argumente darauf hinaus, die Produktivität von Gedanken innerhalb eines solchen sprachähnlichen Repräsentationssystems erklären zu können. Inhaltlich hängen die Argumente in der Weise zusammen, dass die Produktivität durch die Kompositionalität und zugleich die Systematizität von Gedanken gewährleistet wird, die allesamt durch ein sprachähnliches Repräsentationssystem ermöglicht werden.

(ii) *Die LoT als Bedingung für die Kompositionalität und Systematizität von Gedanken*

Gedanken seien produktiv, weil sie kompositional und systematisch sind. Die Produktivität ist zunächst die Eigenschaft eines repräsentationalen Systems, unendlich viele syntaktisch und semantisch distinkte Symbole hervorbringen zu können. Die Systematizität ist währenddessen die Eigenschaft, die ein solches System besitzt, damit die Symbole in verschiedenen Kontexten mit dem gleichen semantischen Wert auftreten können. Auch wenn derweil unklar bleibe, was die Kompositionalität ausmacht, ist jedoch für Fodor eindeutig, dass diese für seine Theorie der Semantik von Gedanken unabdingbar ist:

Nobody knows exactly what compositionality demands, but everybody knows why its demands have to be satisfied. (Fodor 2001: 6)

Die Kompositionalität von Gedanken sei nicht unabhängig von der Systematizität und „perhaps they should be viewed as aspects of a single phenomenon“ (Fodor & Pylyshyn 1988: 41): Das, was die Systematizität für die Verarbeitungsprozesse der Gedanken ist, sei die Kompositionalität für die einzelnen Gedanken. Die Systematizität hänge immer von der Struktur oder eben der Kompositionalität der Gedanken ab. Und diese kompositionale Struktur der Gedanken hänge von der Systematik der kombinatorischen Syntax ab. Gedanken seien in der syntaktischen Weise immer systematisch verbunden, als wenn sie aus denselben semantischen Elementen zusammengesetzt, also kompositional, wären. Daher wird auch eine kombinatorische Semantik zusätzlich zur kombinatorischen Syntax postuliert, wobei eine atomare Konstituente eines mentalen Satzes (annäherungsweise) immer denselben semantischen Beitrag zu einem komplexen mentalesischen Ausdruck bietet, in dem sie auftaucht (vgl. Fodor 1987, 1998: 25; Fodor & Pylyshyn 1988: 41 f.; Fodor & McLaughlin 1990). Schließlich sei es wiederum die LoT mit ihrer kombinatorischen Syntax und Semantik, die die notwendigen Bedingungen für die Kompositionalität und Systematizität von Gedanken bereitstellt.

Fodor (1987, 1998: 97 f.) und Fodor & Pylyshyn (1988) argumentieren für die Systematizität von Gedanken im Sinne der Produktivität. Im Rahmen der LoT sei Systematizität (neben der Kompositionalität) eine Bedingung, um überhaupt eine Produktivität von Gedanken zu gestatten: Schon die Fähigkeit, bestimmte Sätze zu produzieren/verstehen, sei durch die Systematizität intrinsisch mit der Fähigkeit verbunden, bestimmte andere Sätze zu produzieren/verstehen (vgl. Fodor 1987: 149, Fodor & Pylyshyn 1988: 37). Als Illustration hierfür führen Fodor & Pylyshyn an, dass kein Sprecher existieren würde, der weiß, wie er in seiner Muttersprache ausdrücken kann, dass John das Mädchen liebt, aber nicht, dass das

Mädchen John liebt (oder auch zu Umformungen anderer Sätze mit beliebiger n-stelliger Relation fähig ist; vgl. Aydede 2010):

What does it mean to say that thought is systematic? Well, just as you don't find people who can understand the sentence 'John loves the girl' but not the sentence 'the girl loves John,' so too you don't find people who can *think the thought* that John loves the girl but can't think the thought that the girl loves John. (Fodor & Pylyshyn 1988: 39)

Die Möglichkeit dieser Umformungen besteht in der Systematizität (und weiter auch der Kompositionalität) von Gedanken. Wenn ein Muttersprachler fähig ist, bestimmte Sätze in seiner Sprache zu produzieren/verstehen, dann soll es entsprechend eine Struktur wie die der LoT geben, um auch eine andere Gruppe von Sätzen zu produzieren/verstehen. Die Fähigkeit, einen Satz zu verstehen, sei also mit der Fähigkeit verbunden, unendlich viele andere Sätze zu verstehen. Dies ist verbunden mit der Fähigkeit, einen Gedanken zu denken, die mit der Fähigkeit verbunden ist, auch prinzipiell unendlich viele andere Gedanken denken zu können. Um einen Satz zu verstehen, ist es letztlich, wie gesagt, nur nötig, den Gedanken zu denken und die Proposition zu hegen, die er ausdrückt.

Fodor & Pylyshyn (1988: 37 ff.) verdeutlichen den Punkt, dass eine sprachähnliche Struktur die Systematizität ermöglicht, durch einen Vergleich zwischen dem gewöhnlichen Erlernen einer Sprache und dem einer Fremdsprache durch einen Sprachführer¹¹, in dem Übersetzungen von Sätzen, wie beispielsweise „Wo bitte geht es zum Bahnhof?“ stehen. Mit dem Sprachführer werden jedoch nur bruchstückhafte linguistische Fähigkeiten vermittelt, während das linguistische Wissen eines Muttersprachlers systematisch sei. Schon die Grammatik gebe dem Sprecher vor, wie man die Wörter seines Vokabulars systematisch in Sätzen kombinieren kann. Dabei seien Sätze aber nicht atomar wie die Phrasen in dem Sprachführer, sondern würden durch die Grammatik und das Vokabular eine umfassende syntaktisch konstituierte Struktur besitzen, die ein Muttersprachler wiederum kennt. Der Fremdsprachler kann jedoch nicht ohne weiteres durch die Phrasen in seinem Buch neue Sätze bilden, da er die Struktur und zugrundeliegenden Regeln nicht kennt. Nur unter der Bedingung, dass Sätze eine syntaktisch konstituierte Struktur haben, nach der man einzelne Sätze aus bestimmten Wörtern bildet, kann man systematisch viele andere bilden (vgl. auch Fodor 1987: 149). Und eben diese Bedingung für die Systematizität soll die LoT aufgrund ihrer kombinatorischen Syntax (und Semantik) bieten.

Die Systematizität (und Kompositionalität) sei auch grundlegend für die Fähigkeit, bestimmte Inferenzen zu machen. Diese sei weiter intrinsisch mit der Fähigkeit

¹¹ Schneider (2009: 284) argumentiert mit diesem Sprachführer-Beispiel direkt für die Produktivität von Gedanken.

verbunden, bestimmte andere Inferenzen zu ziehen. Die Annahme der Systematizität bei inferentiellen Denkprozessen basiert zunächst auf der Beobachtung, dass wenn ein Mensch fähig ist, eine bestimmte Inferenz zu machen, er dann auch fähig ist, viele andere zu machen, die ähnlich organisiert sind, wie Fodor & Pylyshyn anmerken:

You don't, for example, get minds that are prepared to infer John went to the Store from John and Mary and Susan and Sally went to the store and from John and Mary went to the store but not from John and Mary and Susan went to the store. Given a notion of logical syntax [...] it is a truism that you don't get such minds. (Fodor & Pylyshyn 1988: 48)

Jemanden, der p aus $p \ \& \ q$ ableiten kann, aber nicht r aus $p \ \& \ q \ \& \ r$, wird man nach Fodor & Pylyshyn ebenfalls nicht finden. Und jedes Beispiel für inferentielle Kohärenz könne wie in den anderen bereits genannten Fällen am besten mit einem sprachähnlich strukturierten Repräsentationsmedium wie der LoT erklärt werden. Auch die Kohärenz des Denkens hänge von der logischen Form und Struktur ab, die durch die kombinatorische Syntax der LoT gegeben ist. Die kombinatorische Syntax legt dabei die logische Form für jeden wohlgeformten Ausdruck fest, denn, wenn ein System mentaler Repräsentationen eine kombinatorische Syntax besitzt, dann gäbe es auch eine Menge von Bildungsregeln, nach der die Konstruktion von wohlgeformten Ausdrücken abläuft. Die Regeln bestimmen weiter die syntaktische oder formale Struktur von komplexen mentalen Repräsentationen innerhalb von Inferenzen. Die Möglichkeit von systematischen Inferenzen sei also dadurch gewährleistet, dass mentale Operationen an den Repräsentationen abhängig von ihrer Syntax sind. So sei auch die inferentielle Kohärenz im menschlichen Denken am besten durch ein sprachähnlich strukturiertes Repräsentationsmedium wie die LoT zu erklären.

(iii) *Die LoT als Bedingung des kognitionswissenschaftlichen Erfolgs*

Für ein sehr allgemeines, die anderen Argumente beinhaltendes Argument für die Existenz einer LoT verweist Fodor auf den bisherigen Erfolg der klassischen Kognitionswissenschaft¹²: Unsere derzeit besten kognitionswissenschaftlichen Theorien, also die KTG und die RTG, ließen eine Struktur annehmen, die ein repräsentationales und komputationales Medium wie eine LoT benötigen, damit sie wahr sein können. Da sich aus diesen Theorien so erfolgreiche Modelle zur Erklärung von vielfältigsten Aspekten von Kognition ableiten lassen, sind diese Theorien

¹² Aydede (2010) nennt dieses Argument „Argument from Contemporary Cognitive Psychology“, während Schneider (2009: 285) Fodors inhaltlichen Punkt dem Argument der inferentiellen Kohärenz zuordnet (s.u.).

auch zutreffend. Wie schon oben erwähnt, lässt sich beispielsweise Spracherwerb anhand eines solchen Modells als Hypothesenbildung, -prüfung und -bestätigung auf Basis der KTG und RTG erklären. Ebenso könnte man auf dieselbe Weise auch den Erwerb von Begriffen erklären, die Bildung von Wahrnehmungsurteilen oder die Fähigkeit des Schlussfolgerns.

Entsprechend unserer besten verfügbaren Modelle auf Basis der KTG und der RTG werden kognitive Prozesse also als komputationale Verarbeitung von symbolischen Repräsentationen verstanden. Wenn diese psychologischen Modelle von durch Repräsentationen festgelegten komputationalen Prozessen zutreffend sind – und gerade das wird von Fodor angenommen –, muss es eine LoT geben, in der sie festgelegt sind (vgl. Fodor 1975: 27); also sei die LoTH wahr.

2.2.4 Fodors Modularity of Mind-Hypothese

Fodors *The Modularity of Mind* (1983) hatte weitreichenden Einfluss auf das Denken über den Aufbau des Geistes und das Bild einer kognitiven Architektur. Die Grundidee bei der Annahme einer Modularität des Geistes ist, dass verschiedene kognitive Prozesse getrennt und voneinander isoliert ablaufen. Die einzelnen Prozesse werden relativ unabhängigen Subsystemen, Domänen oder eben Modulen zugeordnet, die von nichts anderem beeinflusst werden, als das, was ihnen als spezifischer Input gegeben ist. Dass unterschiedliche kognitive Subsysteme existieren, ist aber schon eine ältere Idee, die besonders populär im 19. Jahrhundert war und sich beispielsweise deutlich in der Phrenologie von Gall wiederfinden lässt. Gall ging aber nicht nur davon aus, dass voneinander unabhängige Domänen für verschiedene kognitive Fähigkeiten zuständig sind, sondern auch, dass sich diese physisch, und speziell kranial, lokalisieren lassen. Die Ausprägung eines psychischen Vermögens würde, so Gall, mit der Größe der Erhebungen am Kopf an einem entsprechenden Ort korrelieren. Abseits der Annahme, dass die Charaktereigenschaften und Fähigkeiten einer Person etwa an der Schädelform ablesbar sind oder kognitive Fähigkeiten hinsichtlich ihrer Verortung im Gehirn individuiert werden können, bietet die Phrenologie Galls hinsichtlich der Konzeption einer Domänenspezifität von kognitiven Fähigkeiten den Ausgangspunkt für Fodors Modularitätsthese. Fodor übernimmt von Gall die Annahme, dass sich kognitive Prozesse nicht als Interaktion von mehreren Vermögen gleichzeitig verstehen lassen. Gegenüber den sogenannten „horizontalen“ Vermögen, bei denen funktional unterscheidbare kognitive Prozesse mehrere Domänen durchlaufen, seien „vertikale“ Vermögen anzunehmen. Als horizontale Vermögen werden beispielsweise Erinnerung, Vorstellung, Aufmerksamkeit, Empfindungsvermögen und Perzeption gesehen (vgl. Fodor 1983: 11). Hingegen seien z.B. im auditorischen Bereich das

diskriminatorische Erkennen von Sprachlauten, musikalischen Töne oder Umweltgeräuschen verschiedenen vertikalen Vermögen zuzuordnen. Diese Vermögen betreffen zwar alle das Hören, stellen aber für Fodor jeweils einzelne kognitive Fähigkeiten dar, da sie unterschiedliche Funktionen erfüllen. Einige zentrale Eigenschaften, die Gall diesen vertikalen Vermögen zuschreibt, werden später von Fodor für das übernommen, was er „Modul“ nennt:

We can distinguish four major ingredients in Gall's notion of a fundamental power: vertical faculties are *domain specific*, they are *genetically determined*, they are associated with *distinct neural structures*, and—to introduce a new point—they are *computationally autonomous*. The relevant consideration about computational autonomy is that Gall's fundamental powers do not share—and hence they do not compete for—such horizontal resources as memory, attention, intelligence, judgment or whatever. This view of vertical faculties as not merely distinct in the functions they perform, but also relatively independent in the performance of their functions, will be important later when we turn to consider the notion of a cognitive module. (Fodor 1983: 21; Hervorhebung im Original)

Es gibt zwar Unterschiede hinsichtlich dessen, welche Vermögen Fodor und Gall zu den horizontalen und welche sie zu den vertikalen zählen, aber Fodor hält für seine Definition von Modul an den wesentlichen Eigenschaften der Gallschen Charakterisierung von vertikalen Vermögen fest:

Roughly, modular cognitive systems are *domain specific*, *innately specified*, *hard-wired*, *autonomous*, and *not assembled*. Since modular systems are domain specific computational mechanisms, it follows that they are species of vertical faculties. (Fodor 1983:37; Hervorhebung A.M.W.)

Nach Fodor sind wie bei Gall die wichtigsten Eigenschaften von Modulen, dass sie domänenspezifisch, angeboren, (neuronal) festgelegt und eigenständig sind, während sie gerade kein gemeinsames Gefüge bilden. Um das, was ein Modul ist, weiter zu spezifizieren, führt er eine funktionale Taxonomie ein, bei der zwischen Transduktoren (*transducers*), Input-Systemen (*input systems*) und zentralen Verarbeitungsinstanzen (*central processors*) unterschieden wird. Die Transduktoren sollen sich am nächsten zu den Sinnesorganen befinden und die sensorisch erworbenen Aspekte der Außenwelt für die Input-Systeme in Repräsentationen übersetzen. Der Außenwelt könne dann durch die Input-Systeme eine Menge von Repräsentationen zugeordnet werden, wie z.B. ein Aktivierungsmuster auf der Retina. Mit Fodors Worten beschrieben:

Whereas transducer outputs are most naturally interpreted as specifying the distribution of stimulations at the 'surface' (as it were) of the organism, the input systems deliver representations that are most naturally interpreted as characterizing the arrangement of *things in the world*. [...] Specifically, the inferences at issue have as their 'premises' transduced representations of proximal stimulus configurations, and as their 'conclusions' representations of the character and distribution of distal objects" (Fodor 1983: 42; Hervorhebung im Original).

Es bedarf laut Fodor zudem einer weiteren Übersetzung, die durch die Input-Systeme vollzogen wird. Während die Transduktoren die Verteilung der Stimuli an

der sensorischen Oberfläche eines Organismus spezifizieren, liefern die Input-Systeme Repräsentationen der Dinge in der Welt und dienen dazu, „to so represent the world as to make it accessible to thought“ (Fodor 1983: 40). Anders als die Transduktoren würden die Input-Systeme für die Möglichkeit von nicht-demonstrativen Inferenzen sorgen, die über die reine Information des Stimulus hinausgehen und beispielsweise Hypothesenbildung ermöglichen, wie im Fall der Perzeption über die Umgebung. Die höchste Ebene der zentralen Verarbeitungsinstanzen, die für allgemeine Inferenzen, Schlussfolgerungen und Überzeugungsbildung verantwortlich sind, ist nach Fodor jedoch selbst vermutlich nicht modular (vgl. Fodor 1983: 101ff.).¹³

Die Input-Systeme, auf die sich der hier interessante Ausdruck „Modul“ bezieht¹⁴, haben bestimmte Eigenschaften, die zusammen einen bestimmten Typ von System charakterisieren. Sobald ein kognitives System die meisten dieser Eigenschaften in einem nennenswerten Ausmaß besitzt, sei es, so Fodor, modular „to some interesting extent“ (Fodor 1983: 37). Einige Eigenschaften der Module, wie z.B. die informationelle Einkapselung als eine wesentliche Eigenschaft, werden dabei aber als explanatorisch wichtiger angesehen als andere. Fodor führt insgesamt neun mehr oder weniger voneinander unabhängige, typische, aber nicht jeweils notwendige Merkmale an (Fodor 1983: 36f. und 47ff.), die mittlerweile einschlägig in der relevanten Literatur geworden sind (für die folgende Aufzählung siehe Prinz 2006):

- (1) Bereichs- oder Domänenspezifität (*domain specificity*)
- (2) Informationelle Einkapselung (*informational encapsulation*)
- (3) Bindung an eine festgelegte neuronale Architektur (*association with fixed neural architecture*)
- (4) Beschränkte Zugriffsmöglichkeit (*limited central access*)

¹³ Gardner (1983) schlägt hingegen vor, dass auch höherstufige kognitive Prozesse und intellektuelle Fähigkeiten modular sind. Auch anderen Ansätzen zufolge sind höherstufige Fähigkeiten modular organisiert (z. B. Lerdahl & Jackendoff 1983) für musikalische Kognition oder Scholl & Leslie (1999) für soziale Kognition und die Annahme eines „Theory of Mind“-Moduls zur Zuschreibung mentaler Zustände Anderer. Nach Carruthers (2006) ist der Geist selbst durch und durch modular und somit auch alle höherstufige Systeme, die für Gedanken zuständig sind. Eine extremere Sicht einer massiven Modularität wird von beispielsweise Sperber (1994) (siehe auch Cosmides & Tooby 1992; Pinker 1997; Samuels 2006) vertreten, wobei eine große Vielfalt von Modulen die Struktur der kognitiven Architektur bilden. Solche Ansichten sind häufig auch an evolutionäre Ansätze gekoppelt, wobei von ubiquitären Modulen ausgegangen wird, bei denen die Domänen der Verarbeitung weitaus abstrakter sind als das Sehen (vgl. z. B. das von Cosmides & Tooby 1992 vorgeschlagene „cheater-detection module“). Fodor aber behauptet, dass höherstufige Kognition ein allgemein holistisches und nicht-modulares System sei.

¹⁴ Auch die Transduktoren seien aufgrund der Anatomie und Funktionsweise der Sinnesorgane in einem trivialen Sinne modular (vgl. Fodor 1983: 48).

- (5) Verbindliche Arbeitsweise (*mandatory operation*)
- (6) Flacher Output (*shallow output*)
- (7) Schnelle Verarbeitung (*fast operation*)
- (8) Charakteristische Entwicklungsmuster (*fixed ontogeny of pace and sequencing*)
- (9) Charakteristische und spezifische Schädigungsmuster (*characteristic and specific breakdown patterns*)

(1) Ein System ist domänenspezifisch, wenn es bei der Informationsverarbeitung auf jeweils bestimmte Gegenstandsbereiche bzw. Aspekte von Gegenstandsbereichen spezialisiert ist. Somit ist es aber zugleich in der Weise eingeschränkt, da es nur eine bestimmte Klasse von Objekten und Eigenschaften gibt, über die es Informationen verarbeiten kann. Der Grad der Domänenspezifität kann währenddessen als Funktion der Reichweite des Inputs, auf den ein System ausgerichtet ist, verstanden werden, wobei die Größe der Reichweite den informationellen Bereich eines Systems bestimmt (vgl. auch Carruthers 2006: 3–7) – oder wie Fodor es formuliert:

[D]omain specificity has to do with the range of questions for which a device provides answers (the range of inputs for which it computes analyses). (Fodor 1983: 103)

Je enger die Reichweite der Inputs ist, die ein System verarbeiten kann, desto enger ist die Reichweite der Probleme, die das System lösen kann. Und je enger die Reichweite dieser Probleme ist, desto domänenspezifischer ist die modulare Einheit. Domänenspezifische Mechanismen seien beispielsweise Systeme für Farbwahrnehmung, visuelle Analyse von Formen, Syntaxanalyse von Sätzen oder Stimm- und Gesichtserkennung (vgl. Fodor 1983: 47). Diese seien aber nicht direkt mit perzeptuellen und linguistischen Fähigkeiten vergleichbar, da sie feinkörniger sind. Fodor spricht zudem von „content domains“ (1983: 36), die er aber von den verschiedenen Sinnesmodalitäten oder vertikalen Vermögen im Sinne Galls unterscheiden möchte (vgl. auch Fodor 2001: 58 ff.).

(2) Ein kognitives System ist informationell eingekapselt, da ein Modul nur für eine bestimmte Menge von Einflüssen zuständig ist, die nicht woanders innerhalb des Systems verarbeitet oder gespeichert werden können.¹⁵ Die Einkapselung ist im Fall der spezifischen Zuständigkeit eine Folge der funktionalen Trennung der Domänen. Diese Eigenschaft lässt sich am besten mit dem illustrieren, was Pylyshyn (1980; 1984) „kognitive Impenetrabilität“ nennt. Für den Fall von Perzeption

¹⁵ Aber ein System kann in der Hinsicht eingekapselt sein, ohne gänzlich eingekapselt zu sein. Beispielsweise ist auditorische Sprachwahrnehmung gegenüber Überzeugungen eingekapselt, aber nicht in Bezug auf Sehen, wie man am Beispiel des McGurk-Effektes zeigen kann (vgl. McGurk & McDonald 1976).

bedeutet dies, dass der Output eines perzeptuellen Systems nicht empfänglich ist für das, was der Wahrnehmende annimmt oder wünscht (vgl. Fodor 1983: 68). Klassische Beispiele dieser Eigenschaft stammen aus Studien mit visuellen Täuschungen, welche typischerweise als Täuschungen andauern, selbst wenn der Betrachter explizit über den Charakter des Stimulus und der Täuschung informiert wurde. So sehen zum Beispiel die zwei Linien in der Müller-Lyer-Illusion weiterhin so aus, als wenn sie unterschiedlich lang wären, selbst wenn man vom Gegenteil überzeugt ist. Dass ein Prozess kognitiv impenetrabel ist, meint also, dass bestimmte Informationen aufgrund der Einkapselung durch andere kognitive und perzeptuelle Prozesse nicht beeinflussbar sind, selbst wenn dies für eine bestimmte Aufgabe vielleicht relevant wäre.¹⁶

(3) Die Bindung an eine festgelegte neuronale Architektur ist als ein Nebeneffekt der Einkapselung zu verstehen. Die Module seien bereits ontogenetisch festgelegt und entwickeln sich in einer vorhersehbarer Art und Weise. Das heißt auch, dass sie angeboren sind und sich nicht erst durch Erfahrung herausbilden. Kognitive Vermögen seien zwar in erster Linie funktional und nicht physiologisch zu individualisieren, aber es gäbe dennoch festgelegte Verbindungen bei bevorzugten Abläufen der Informationsverarbeitung.

(4) Die beschränkte Zugriffsmöglichkeit bezieht sich auf den Ausgang des Informationsfluss, während die Einkapselung die Begrenzung des Eingangs der Informationen zu einem Mechanismus bedingt. Genauer gesagt, ein System hat nur eine beschränkte Zugriffsmöglichkeit in dem Sinne, dass Repräsentationen auf der Ebene, die komputational vor der Produktion eines Outputs liegt, für die Introspektion nicht zugänglich sind. Erst die Outputs eines solchen Systems können beispielsweise eine besondere phänomenale Dimension aufweisen, ihre Vorstufen aber nicht. So beginne auch Sprachverstehen mit verschiedenen Repräsentationen eines Stimulus, wie phonologischer, lexikalischer, syntaktischer Art etc., die sukzessive verarbeitet werden, aber erst als Bedeutung des Gesagten bewusst zugänglich sind (vgl. Fodor 1983: 55 ff.).

(5) Die Operationen eines kognitiven Systems sind verbindlich in dem Sinne, dass sie nicht bewusst kontrolliert werden. Sie werden durch die relevanten Stimuli ausgelöst und automatisch abschließend verarbeitet. So können nach Fodor z.B.

¹⁶ Kognitive Impenetrabilität ist ein Fall der Einkapselung bezogen auf die Informationen, die im Gedächtnis als beispielsweise eine Form von Überzeugung oder Nützlichkeit gespeichert werden. Und der McGurk-Effekt (vgl. McGurk & McDonald 1976) scheint zugleich zu zeigen, dass der Prozess des Erkennens von relativ niedrigstufigen Eigenschaften der Sprache wie von Phonemen penetrabel durch Informationen aus der visuellen Modalität ist, da Gesichtsbewegungen gehörte Phoneme beeinflussen.

englische Muttersprachler nicht nur die bloßen Geräusche des gesprochenen Englisch hören, sondern sie hören diese direkt als Englisch. Entsprechend sei es unmöglich, einen dreidimensionalen Bereich von Objekten im Raum einfach als Nebeneinander von farblichen zweidimensionalen Flächen zu sehen (vgl. Fodor 1983: 52 f.). Auch Marr (1982; s.o.) verweist bereits bei seiner Analyse der visuellen Extrahierung von Tiefeninformation aus stereoskopischen Bildern auf bestimmte Inferenzen, die in einem visuellen System aufgrund von Regeln gemacht werden, zu denen das Subjekt weder bewussten Zugang hat, noch fähig ist, diese zu beeinflussen.

(6) Der Output wird in einem modularen System relativ ‚flach‘ oder komputational möglichst wenig aufwendig gehalten. Einerseits ist entscheidend, wie viel Komputation benötigt wird, um einen Output zu produzieren. Andererseits ist es von Bedeutung, wie spezifisch der informationelle Gehalt ist; ‚wenig aufwendig‘ ist daher als ‚informationell allgemein‘ zu verstehen (vgl. Fodor 1983: 87). Diese beiden Eigenschaften hängen in der Weise zusammen, dass der Output mit einem spezifischeren Gehalt für ein System aufwendiger zu produzieren ist. Zusammengefasst, je mehr ein System informationell eingekapselt ist, desto schneller und weniger aufwendig ist die Komputation. Dies bedeutet aber auch, dass es dadurch kognitiv unkontrollierter bzw. weniger penetraibel ist (vgl. auch Robbins 2010).

(7) Geschwindigkeit hängt von der Effizienz der Verarbeitung ab, die mit der Einkapselung in der Weise korreliert, als dass die Einkapselung die informationelle Belastung eines Systems reduzieren soll. Dieser Punkt muss aber nicht weiter ausbuchstabiert werden, wenn ein Beispiel dienen kann: Bei der experimentellen Technik des sogenannten „speech shadowing“ wiederholen Subjekte Sprache unmittelbar nachdem sie diese gehört haben. Die Reaktionszeit zwischen dem Hören und dem Aussprechen beträgt dabei 254 ms (oder auch nur 150 ms vgl. Marslen-Wilson 1973; bei Fodor 250 ms). Während ein Subjekt die Wörter nur wiederholt, würden automatisch aber auch ihre Syntax und Semantik verarbeitet. Da die Rate der Silben der normalen Sprache bei etwa 4 Silben pro Sekunde liegt, deutet dies darauf hin, dass man den Stimulus in der Dauer einer Silbenlänge verarbeitet (vgl. Fodor 1983: 61 ff.).

(8) Die Angeborenheit charakteristischer Entwicklungsmuster wird als die Eigenschaft verstanden, die sich entwickelt „according to specific, endogenously determined patterns under the impact of environmental releasers“ (Fodor 1983: 100). Fodor orientiert sich an einer Reihe von Studien der ontogenetischen Entwicklung beim Spracherwerb und visuellen Fähigkeiten bei Kleinkindern. Die Befunde sind mit seiner Behauptung kompatibel, dass die Entwicklung von Input-Systemen endogen bestimmt ist. Robbins (2010) führt für den Spracherwerb einen neueren Befund an, dass unabhängig vom Kulturkreis bei allen normalen Individuen mehr

oder weniger derselben Ablauf vorliegt: einzelne Wörter werden ab 12 Monaten gelernt, telegraphisches Sprechen ab 18 Monaten, komplexe Grammatik ab 24 Monaten usw. (vgl. dazu Stromswold 1999).

(9) Dass ein System selektiv beeinträchtigt werden kann, heißt für Fodor, dass es beschädigt, behindert oder gestört werden kann und dies nur eine kleine oder gar keine Wirkung auf die Operationen in anderen Systemen hat. Selektive kognitive Beeinträchtigungen dieser Art würden häufig als eine Konsequenz von jeweils bestimmten Gehirnläsionen beobachtet (vgl. Fodor 1983: 99 f.). So gäbe es Gehirnläsionsstudien, die die Annahme stützen, dass lokale Läsionen Sprachfähigkeiten beeinträchtigen, aber keine anderen Aspekte der Kognition (vgl. z.B. Caramazza et al. 1983). Auch die Fähigkeiten beim Erlernen von Sprache scheinen zum Beispiel relativ unabhängig von allgemeinen Problemlösungsfähigkeiten zu sein (vgl. z.B. Karmiloff-Smith et al. 1997).¹⁷

2.3 Zusammenfassung

Entsprechend der komputationalen Theorie des Geistes funktioniert Kognition analog zu einem Computer, der Symbole verarbeitet. Bisher haben wir verschiedene klassische Konzeptionen und Anwendungsbereiche kennengelernt, in denen diese Idee zum Ausdruck kommt. Die kognitiven Prozesse werden dabei allgemein als eine Form von komputationaler Symbolmanipulation verstanden, die nach bestimmten Regeln abläuft und schließlich dazu führt, dass beispielsweise Lösungen für Probleme gefunden oder Schlussfolgerungen abgeleitet werden. Die regelgeleitete Verarbeitung von Symbolen basiert währenddessen auf den strukturellen Eigenschaften der einzelnen repräsentationalen Elemente und ihren physischen Relationen. Dieser Grundgedanke wurde von Fodor weitergeführt, um auch die Bedeutung von Gedanken zu erklären. Gedanken mit propositionaler Struktur würden sich aus komplexen symbolischen Repräsentationen zusammensetzen, die bestimmte semantische und syntaktische Eigenschaften besitzen und

¹⁷ Standardbeispiele für selektive Beeinträchtigungen sind im Fall von Perzeption Prosopagnosie (beeinträchtigte Gesichtserkennung), Achromatopsie (vollständige Farbenblindheit) und Akinetopsie (Bewegungsblindheit). Weitere Beispiele sind im Fall von Sprachverstehen Agrammatismus (Verlust der komplexen Syntax), Jargonaphasie (Verlust der komplexen Semantik), Alexie (Verlust von Objektwörtern) und Dyslexie (beeinträchtigtes Lesen und Schreiben) (für diese Aufzählung siehe Robbins 2010). Dass jede dieser Störungen bei andererseits normalen Individuen gefunden wurde, deutet daraufhin, dass die verlorenen Fähigkeiten funktional von anderen isolierte Mechanismen betreffen. Dies als Belege für die These einer Modularität anzuführen, ist jedoch eher kontrovers: Farah (1994) stellt beispielsweise heraus, dass Läsions- und Defizitstudien nicht die neurologische Lokalisierbarkeit von Sprachfunktionen stützen, sondern zurückweisen.

deren Verarbeitung formalen Regeln gehorcht. Schließlich würden die strukturellen Eigenschaften der Repräsentationen eine Syntax bilden und auch für eine entsprechende Semantik verantwortlich sein. Dieses Repräsentationssystem würde weiter die interne Basis für Fähigkeiten, wie den Erwerb von Sprache, das Gedächtnis und den Verstand im Allgemeinen, bilden. Fodors sprachähnliche Auffassung von Kognition, dass Gedanken aufgrund ihrer symbolischen Struktur eine Syntax mitsamt einer Semantik aufweisen, ist dabei wie bei Newell & Simon verbunden mit einem speziellen Verständnis von mentalen Repräsentationen als Symbolen.

Der repräsentationalen Theorie des Geistes im Sinne der klassischen Auffassung folgend ist das Denken eines bestimmten Gedankens ein Vorkommen einer bestimmten komplexen Repräsentation. Die Fähigkeit, eine bestimmte Repräsentation zu haben, die die relevante Proposition ausdrückt, ist systematisch mit der Fähigkeit verbunden, auch bestimmte andere zu haben. Die beteiligten Repräsentationen werden nach Fodor in einem sprachähnlichen Repräsentationsmedium wie der LoT verarbeitet und der Geist beinhaltet dabei eine Reihe von Symbolen, die entsprechend grammatikalischer Prinzipien wie Wörter zu mentalen Sätzen kombiniert werden. Dabei geht es aber nur um die mentalen Zustände, die explizit und bewusst sind, also nicht um solche, die nicht bewusst sind, weil sie auf einer zu niedrigen Ebene der Verarbeitung auftreten. Marr entwickelte entsprechend eine Beschreibung der visuellen Prozesse, bei der drei unterschiedliche Ebenen berücksichtigt werden: die Ebene der Implementation, des Algorithmus und der Komputation. Nur unter Berücksichtigung dieser Ebenen sei zu beschreiben, wie der rein sensorische und niedrigstufige Input zu einem höherstufigen perzeptuellen Output führt. Zudem sei erst auf der höchsten Ebene der komputationalen Verarbeitung eine Untersuchung der Repräsentationen und des dazugehörigen Algorithmus möglich. Auch in anderen klassisch angelegten Arbeiten wird zwischen drei Ebenen unterschieden, die hinsichtlich ihrer Merkmale denen in Marrs Beschreibung gleichen. Eine solche Dreiteilung findet sich beispielsweise bei Pylyshyn (1984), der zwischen einer semantischen, syntaktischen und physischen Ebene trennt:

[T]here are actually two distinct levels above the physical or neurophysiological level—a representational or semantical level and a symbol-processing level. (Pylyshyn 1984: 24)

Die Ebene der Symbolverarbeitung entspricht bei Marr der algorithmischen Ebene und die der Semantik der komputationalen Ebene, während die Implementationsbasis eine physische oder neurophysiologische Ebene darstellt. Auch Glass, Holyoak & Santa (1979) unterscheiden auf eine ähnliche Weise die Ebene des Gehalts, der Form und des Mediums, wobei es hier neben den rein terminologischen Unterschieden ebenfalls keine besonderen Abweichungen hinsichtlich der

Gegenstandsbereiche gibt. Diese Ähnlichkeit stellt Pylyshyn so dar, dass den drei Ebenen jeweils unterschiedliche Prinzipien zugrunde liegen:

[T]he principle generalizations covering behavior occur at three autonomous levels of description, each conforming to different principles. These principles are referred to as the biological (or physical) level, the symbolic (or syntactic or sometimes the functional) level, and the semantic (or intentional) level. (Pylyshyn 1984: 259)

Die biologische (oder physische) Ebene korrespondiert hier mit Marrs Ebene der Implementierung, die symbolische (oder syntaktische) Ebene mit der der algorithmischen Verarbeitung der Repräsentationen und die semantische (oder intentionale) Ebene mit der der Komputation. Diese Prinzipien lassen sie auch in Fodors LoT-Konzeption wiederfinden. Die Aufteilung in drei Ebenen ist nicht nur charakteristisch für die klassische Sicht auf die Weise der allgemeinen kognitiven Verarbeitung, sondern auch auf die Gedanken, die nicht schon auf der biophysiologicalen, physischen Ebene der Implementation oder der symbolisch-syntaktischen Ebene der algorithmischen Verarbeitung zu erklären sind, sondern erst auf der höchsten Ebene der Komputation, der Ebene der Semantik und Intentionalität.

Weiter haben wir mit Fodors These über die Modularität des Geistes ein besonderes Bild einer kognitiven Architektur präsentiert bekommen. Hier finden wir ebenfalls eine Dreiteilung hinsichtlich seiner funktionalen Trennung zwischen Transduktoren, Input-Systemen und zentralen Verarbeitungsinstanzen. Die modular geprägten, voneinander unabhängig arbeitenden Input-Systeme sind jeweils für verschiedene kognitive Aufgaben zuständig und nutzen dabei auch unterschiedliche Repräsentationen, die schließlich auf der höchsten Ebene die Intentionalität eines kognitiven Systems mit einem breiten begrifflichen Spektrum gewährleisten. Diese Intentionalität ist währenddessen nicht mehr ausschließlich durch die verschiedenen Sinnesmodalitäten bestimmt, sondern nach unterschiedlichen kognitiven Fähigkeiten differenziert, sodass beispielsweise auf der Basis von visuellen Stimuli unterschiedliche diskriminatorische Fähigkeiten, wie Objekt- oder Gesichtserkennung, möglich werden und von auditorischen Stimuli ausgehend Sprachverstehen oder die Unterscheidung von Geräuschen und Musik.

2.4 Probleme der klassischen Kognitionswissenschaft

2.4.1 Konnektionismus

Die KTG und die RTG waren zwar spätestens in den 1980er Jahren die populärsten Theorien in der Kognitionswissenschaft, aber sicherlich nicht ohne auch Alternativen gegenüberzustehen und immer wieder erneuter Kritik ausgesetzt zu

sein. Eine besondere Debatte entwickelte sich durch das Aufkommen des Konnektionismus, der auf gänzlich anderen Annahmen zur Erklärung von Kognition beruht, mit dem sich u.a. aber viele neue und erfolgreiche Möglichkeiten in der Forschung zur KI boten. Zugleich eröffnete sich durch den Konnektionismus auch eine neue Sichtweise auf die Funktionsweise und die Architektur des Geistes (vgl. zu den weitreichenden Implikationen und vielen Anwendungsgebieten Ramsey, Stich & Rumelhart 1991). Der Unterschied gegenüber der klassischen Auffassung besteht im Wesentlichen zunächst darin, dass nicht davon ausgegangen wird, dass die kognitive Verarbeitung seriell von einem Input zu einem Output verläuft, sondern über vielfältige Verflechtungen und Aktivierungen in einem neuronalen Netzwerk.

Bereits die klassische Annahme, dass kognitive Prozesse komputational ablaufen und das Gehirn mit der Hardware eines Computers zu vergleichen ist, oder der Geist mit der Software, sei genauer betrachtet nicht besonders naheliegend. Ein Gehirn ist eine Sammlung von Zellen mit Tausenden von Verflechtungen, zusammengesetzt aus vielen Einheiten und ihren Verbindungen, den Neuronen und Synapsen, die zunächst gar keine Ähnlichkeiten zu den Hardwarekomponenten eines Computers aufweist (siehe für diesen Vergleich Shapiro 2011: 41 ff.). Skeptisch gegenüber der klassischen Kognitionswissenschaft präsentieren Smolensky et al. die Grundzüge des Konnektionismus und ihre Präferenz für ein Bild der Informationsverarbeitung von miteinander verbundenen Neuronen und ihren Aktivierungsmustern mit folgenden Worten:

Formal theories of logical reasoning, grammar, and other higher mental faculties compel us to think of the mind as a machine for rule-based manipulation of highly structured arrays of symbols. What we know of the brain compels us to think of human information processing in terms of manipulation of a large unstructured set of numbers, the activity levels of interconnected neurons. (Smolensky, Legendre & Miyata 1992: 382)

Weiter solle man davon ausgehen, dass Gedanken auf flexible Muster neuronaler Aktivierungen zurückgehen, die sich numerisch bestimmen lassen, jedoch keine starren syntaktischen Eigenschaften besitzen und somit auch keine verbindliche Struktur für eine regelgeleitete Verarbeitung bieten. Kognitive Fähigkeiten sollen daher besser durch Modelle von neuronalen Netzen oder Netzwerken erklärt werden können. Vor diesem Hintergrund werden zugleich mehrere Annahmen der klassischen Auffassung in Zweifel gezogen: Erstens, dass die Verarbeitung im Rahmen einer seriellen Input-Output-Relation zu fassen ist, zweitens, dass symbolische Repräsentationen die Elemente der Verarbeitung darstellen und, drittens, dass währenddessen die Implementierungsebene, sei es ein Gehirn oder eine Computerhardware, keine Rolle für die algorithmische und funktionale Beschreibung spielt. Mit der Stichhaltigkeit der konnektionistischen Kritik stehen weiter auch die Annahme einer Systematizität von Gedanken und damit auch die klassische

Erklärung dieser durch ein sprachähnliches und symbolbasiertes Repräsentationsmedium auf dem Spiel.

Ein weiterer Ausgangspunkt der Kritik war die allgemeine Unzufriedenheit mit der klassischen Konzeption der KI, die mittlerweile von Konnektionisten und anderen nicht-klassisch orientierten Vertretern häufig mit dem von Haugeland eingeführten Ausdruck „GOFAI“ (*good old fashioned artificial intelligence*) bezeichnet wird. Beim künstlichen Modellieren kognitiver Leistungen sei es nicht ausreichend, allein die rein internen Systemzustände und eine komputationale Logik zu berücksichtigen, wenn es darum geht, dass ein System flexibles und angepasstes Verhalten generieren bzw. sogar lernfähig sein soll. Bei zahlreichen Anwendungen von konnektionistischen Modellen in Form von künstlichen neuronalen Netzen wurden zusätzlich kein internes Repräsentationsmedium und auch keine dort stattfindende regelgeleitete Symbolmanipulation mehr vorausgesetzt. Dennoch war es möglich, dass Fähigkeiten, wie beispielsweise Gesichts- und Objekterkennung, Lesen und das Bestimmen einfacher grammatikalischer Strukturen, von künstlichen neuronalen Netzen erlernt werden konnten (vgl. dazu Rumelhart & McClelland 1986; Smolensky 1999).

(i) *Neuronale Netze und Aktivierungsmuster*

Neuronale Netze sind zunächst Teile des Gehirns, zusammengesetzt aus einer großen Anzahl von Einheiten, den Neuronen, und den Verbindungen zwischen diesen Einheiten, den Synapsen. Die Einheiten werden in konnektionistischen Modellen typischerweise in drei Klassen unterteilt: Input-Einheiten, die die Informationen empfangen, die später verarbeitet werden, Output-Einheiten, die die Resultate der Verarbeitung anzeigen, und die sogenannten „versteckten“ Einheiten (*hidden units*) zwischen Input und Output. Übertragen auf das menschliche Nervensystem sind die Input-Einheiten sensorische Neuronen, die Output-Einheiten motorische Neuronen, und die versteckten Einheiten analog zu allen anderen Neuronen zu verstehen. Eingehenden Informationen werden schrittweise über die Verbindungen zwischen den Einheiten weitergeleitet, wobei schließlich die Output-Einheiten das Ergebnis anzeigen. Die Informationsverarbeitung wird hierbei bestimmt durch das Muster der Gewichtungen der verschiedenen Verbindungen zwischen den Einheiten. Dieses Muster wird entsprechend eines zugrundeliegenden (Lern-)Algorithmus vorher erlernt (vgl. dazu allgemein Besold & Kühnberger 2013; Garson 2012; McClelland & Rogers 2003).

In der KI wurden verschiedene Modelle künstlicher neuronaler Netze entwickelt, wie z.B. Feed-Forward-Netze mit den genannten drei Ebenen, von den Input-Einheiten über die versteckten Einheiten zu den Output-Einheiten (Rosenberg 1958

bot hierfür die konzeptionelle Grundlage). In einem solchen vorwärts gerichteten Netz produzieren die gleichen Inputs jedes Mal den gleichen Output. Realistischere Modelle für menschliche Kognition ähnlich dem Gehirn beinhalten zudem mehrere Ebenen der versteckten Einheiten, bei denen auch Signale zurück von den späteren höheren Ebenen zu den niedrigeren früheren Ebenen gesendet werden können (vgl. für eine frühe Konzeption siehe Minsky & Papert 1969 und vgl. auch das Backpropagationsmodell von Bryson & Ho 1969, bei dem innerhalb des Algorithmus die Ausgabewerte mit korrekten Ergebnissen verglichen werden, um diese Informationen innerhalb des Netzes zurückzupropagieren, damit durch den Algorithmus die Gewichtung der Verbindungen entsprechend angepasst wird, um den Wert der Fehlerfunktion zu verringern). Bei rekurrenten Netzen, die in den 1980ern aufkamen (vgl. z.B. Hopfield 1982 zu Netzen, deren Einheiten zugleich eine Eingabe- und Ausgabeschicht haben), können gezielte Rückkopplungseffekte mit früheren Ebenen erzeugt werden (vgl. Elman 1990 für die Ergänzung durch sogenannte Kontextzellen als die simpelste und bekannteste Form rekurrenter Netze). Derartige Konzeptionen können beispielsweise dazu dienen, kognitive Fähigkeiten wie das Kurzzeitgedächtnis zu modellieren.

Der konnektionistischen Sicht nach ist kognitive Verarbeitung bei einem Organismus wie dem Menschen als dynamische und graduelle Entwicklung von Aktivität in einem neuronalen Netz zu verstehen, wobei die Aktivierung jeder Einheit von der Verbindungsstärke zu seinen Nachbarn und ihrer jeweiligen Aktivität abhängt. Die sich so ergebenden Aktivierungen bilden ein Muster, das durch die Gewichtung oder Stärke der Vernetzung zwischen den Einheiten bestimmt ist. Im Hintergrund steht hier die Hebb'sche Regel (vgl. Hebb 1949), dass sich die Neuronen miteinander verbinden, die gemeinsam aktiviert werden. Die Gewichtung kann dabei positiv oder negativ sein, wobei eine negative Gewichtung eine Beschränkung der empfangenden Einheiten durch die sendenden Einheiten in Form des fehlenden Inputs ist oder eine Restriktion durch den zugrundeliegenden Algorithmus. Wenn weiter angenommen wird, dass alle Einheiten in einem Gehirn in etwa die gleiche Aktivierungsfunktion „berechnen“, sollen kognitive Leistungen in erster Linie von den Mustern der Gewichtungen zwischen den Einheiten abhängen (vgl. ausführlicher zu dem algorithmischen Kalkül Dawson 1998 und speziell Dawson 2005: 6–8).

Neuronale Netzwerke haben zudem den allgemeinen Vorteil, dass sie sich gut an bestimmte Aufgaben und Probleme anpassen können, da die Verarbeitung nicht zentralen Regeln folgt, die sich in einem klassischen Sinne auf einer algorithmischen Ebene funktional beschreiben lassen. Da die Verarbeitung verteilt und parallel abläuft, können auch die Lösungen bestimmter Aufgaben parallel verarbeitet

werden (das sogenannte *Parallel Distributed Processing*, oder auch *PDP*; vgl. Rumelhart & McClelland 1986; siehe auch van Gelder 1990). Während beispielsweise bei der Informationsverarbeitung durch das Gehirn verzerrter Input oder die Zerstörung von einzelnen Einheiten zwar zur allmählichen Degeneration (*graceful degradation* oder *degeneration*) von Funktionen führt, ist die Reaktion des gesamten Netzes immer noch relativ angemessen, wenn auch sicherlich nicht mehr so akkurat. Wenn nur einzelne Teile des Netzwerkes zerstört oder überlastet werden, sind im Aktivierungsmuster verteilte Repräsentationen an separaten Orten wenig betroffen. In einem Computer führen Verzerrungen oder der Verlust von Schaltkreisen jedoch zu einem vollständigen Versagen einzelner oder auch aller Funktionen. Die neuronalen Netze wie auch ihre künstlichen Umsetzungen in der KI besitzen somit eine gewisse Flexibilität hinsichtlich der Herausforderungen durch die Welt und sollen daher schließlich ein realistischeres und natürlicheres Modell für Kognition bereitstellen.

(ii) *Subsymbolische Muster und Systematizität*

Der Unterschied zwischen der konnektionistischen Annahme neuronaler Netze und der klassischen Auffassung komputationaler Verarbeitung symbolischer Repräsentationen liegt im Wesentlichen darin, dass einerseits als Bedingungen für kognitive Prozesse neuronale Aktivierungsmuster postuliert werden, die über große Teile des Kortex verteilt sind, und andererseits nur lokale oder atomare Repräsentationen mit symbolischen Eigenschaften. Wenn Repräsentationen in Mustern und nicht in individuellen Einheiten wie symbolischen Repräsentationen kodiert sind, ist die Beziehung zwischen den Repräsentationen in den Ähnlichkeiten und Unterschieden zwischen den Mustern kodiert und nicht über die syntaktischen (und semantischen) Eigenschaften der Repräsentationen. Damit ist der Konnektionismus eine neue Sicht gegenüber einer klassischen Auffassung, dass einzelne Neuronen (oder vielleicht auch Bündel von Neuronen) für bestimmte Informationsverarbeitungen zuständig sind. Smolensky (1988) formulierte seinen konnektionistischen Ansatz explizit als eine Gegenposition zum klassischen Bild der Symbolmanipulation, das „symbolische Paradigma“. Sein Ansatz ist mit zwei Thesen verbunden: (1) dass der Prozessor ein sub-symbolisches, konnektionistisches und dynamisches System ist, das keine vollständige, formale und präzise begriffliche Ebene der Beschreibung zulässt, und (2), dass die Festlegung auf eine sub-begriffliche Ebene der Beschreibung als eine direkte Ablehnung der komputationalen Symbolverarbeitung aufzufassen ist:

The name „subsymbolic paradigm“ is intended to suggest cognitive descriptions built up of *constituents* of the symbol used in the symbolic paradigm; these fine-grained constituents might be called *subsymbols*. Entities that are typically represented in the symbolic paradigm by symbols are

typically represented in the subsymbolic paradigm by a large number of subsymbols. Along with this semantic distinction comes a syntactic distinction. Subsymbols are not operated upon a "symbolic manipulation": they participate in numerical—not symbolic—computation. (Smolensky 1988: 3)

Die kognitive Verarbeitung läuft nach Smolensky in einer konnektionistischen Architektur ab, deren Konstituenten die numerischen Aktivitäten der einzelnen Verarbeitungen innerhalb eines Netzwerkes sind. Diese numerischen Aktivitätsmuster sorgen für die sogenannten Subsymbole, die als Konstituenten der Symbole dienen. Nicht einzelne Neuronen (oder Neuronenbündel) fungieren als Symbole, sondern komplexe Gesamtaktivitätsmuster von Neuronen, die die Aufgabe der Symbole im klassischen Sinne schließlich auch besser erfüllen könnten. Die Interaktionen beteiligter Neuronen werden also nicht als symbolisch, sondern als in Abhängigkeit von ihren numerischen Aktivierungswerten stehend gesehen. Smolensky (1991) weist weiter daraufhin, dass verteilte Repräsentationen für komplexe Ausdrücke wie „Mary liebt John“ in einem konnektionistischen Modell so konstruiert werden können, dass sie keine explizite Repräsentationen ihrer Konstituenten „Mary“, „liebt“ und „John“ beinhalten. Dies trifft u.a. direkt die Annahme Fodors, dass eine LoT samt kombinatorischer Syntax die Voraussetzung für menschliche Kognition sei. Und dies ist einer der Unterschiede der Annahme subsymbolischer Repräsentationen gegenüber der klassischen Annahmen wie in der LoTH, bei der von einer kognitiven Verarbeitung als Manipulation von symbolischen und auch bewussten Repräsentationen in Form von Begriffen ausgegangen wird (vgl. allgemein dazu auch Cummins 1989, 1991).

Aufgrund des Fehlens von kausal in komplexen Repräsentationen wirkenden symbolischen Konstituenten muss also nicht notwendigerweise auch eine Systematizität folgen, da durch das Fehlen der expliziten symbolischen Repräsentation auch keine syntaktischen Eigenschaften gegeben sind, die die Verarbeitung gemäß einer kombinatorischen Syntax (und Semantik) bestimmen. Auch wenn konnektionistische Modelle darauf trainiert werden könnten, systematisch zu sein, könnten sie auch darauf trainiert werden „John liebt Mary“ zu verstehen, aber nicht „Mary liebt John“; womit die Annahme von Kompositionalität und Produktivität nicht notwendig erscheint (vgl. auch Aizawa 1997; Chalmers 1993; Gelder, van 1990; Hadley 1997; Matthews 1997). Wenn der Konnektionismus keine Garantie für Systematizität biete, könne aber nicht erklärt werden, warum Systematizität im menschlichen Denken vorliegt, so die Gegenantwort von Fodor & Pylyshyn (1988) und Fodor & McLaughlin (1990). Systematizität existiert in konnektionistischen Architekturen nur rein zufällig, aber nicht notwendig wie in den klassischen Modellen. Durch die Zufälligkeit von Systematizität steht dann auch die Alltagspsychologie und vor-wissenschaftliche Sicht auf Gedanken in Frage, die von Fodor in seiner Konzeption miteinbezogen wurde.

Es gibt aber seit dem Aufkommen des Konnektionismus auch vermittelnde Positionen mit der Annahme, dass das neurale Netz des Gehirns einen symbolischen Prozessor auf höherer und abstrakterer Ebene der Beschreibung implementiert. In diesen sogenannten hybriden Ansätzen werden sowohl neuronale Aktivierungsmuster wie auch symbolischen Repräsentationen postuliert, um auch höherstufige Fähigkeiten wie Schlussfolgern erklären zu können. Für viele andere Konnektionisten liegt währenddessen jedoch der Fehler der klassischen Auffassung immer noch darin, dass die Erklärung der Flexibilität und Effizienz der menschlichen Kognition in Form einer Anpassung an eine unbekannte Umgebung, über die noch keine internen Informationen gegeben sind, ein neues Paradigma der Kognitionswissenschaft benötigt. So sind radikale Konnektionisten der Meinung, dass Symbolverarbeitung gänzlich aus der Kognitionswissenschaft eliminiert werden sollte, während andererseits gemäßigte Konnektionisten ihre Position nicht als Herausforderung der klassischen Sicht sehen.

2.4.2 Harnads *Symbol Grounding-Problem*

Für Harnad (1990) ist durch klassische Erklärungen wie die LoTH noch keine befriedigende Antwort auf die Frage gegeben worden, wie Symbole (und damit auch Gedanken) ihre Bedeutung bekommen, während sie sonst nur in Beziehungen zu anderen bedeutungslosen, rein syntaktisch strukturierten Symbolen stehen. Das bisherige Ausbleiben einer informativen Antwort nennt er das *Symbol Grounding-Problem*. Die Antwort auf die Frage, wie semantischer Gehalt von Gedanken zustande kommt, sei auch nicht durch einen Verweis auf die syntaktische Verarbeitung von Symbolen zu geben, sondern nur durch eine Erklärung, die Auskunft über die sensorische Basis (*ground*) der Repräsentationen gibt.

Den Ausgangspunkt für Harnads Problemdiagnose der klassischen Erklärung bietet Searles (1980) Gedankenexperiment des Chinesisch-Zimmers, das ursprünglich als Kritik gegenüber der Möglichkeit einer starken Form von künstlicher Intelligenz konzipiert war, dass diese den Eigenschaften und Möglichkeiten der menschlichen Intelligenz mindestens gleichkommt. Zudem soll, so Searle, dieses Gedankenexperiment aber auch als ein Argument gegen den Funktionalismus und den Komputationalismus verstanden werden und sich gegen die klassische Annahme richten, der Geist sei ein rein komputationales, symbolmanipulierendes System, wie z.B. die Turing-Maschine (vgl. Searle 1992: 44).

In diesem Gedankenexperiment geht es um eine englischsprachige Person, die die chinesische Sprache nicht beherrscht. Diese Person, so stelle man es sich vor, befindet sich in einem Zimmer mit einem Schlitz in der Wand für einen Kontakt mit der Welt außerhalb. Durch diesen Schlitz kommen Zettel in das Zimmer, die mit

chinesischen Schriftzeichen beschriftet sind. In dem Zimmer sind des weiteren Anleitungen in englischer Sprache, die angeben, welche chinesischen Symbole für eine entsprechende Antwort auf einen Zettel geschrieben werden sollen, um dann wieder durch den Schlitz nach draußen gegeben zu werden. Diese Antworten kann die Person nach einiger Zeit mit Hilfe der Anleitungen recht schnell geben und für einen Chinesisch-Sprecher außerhalb scheint es dann so, als wenn die Person im Zimmer Chinesisch versteht. Nach Searles Intuition ist ein Verständnis der Schriftzeichen auf den Zetteln durch die Person in dem Zimmer jedoch nicht gegeben.

Dieses Gedankenexperiment ist analog zum Turing-Test (vgl. Turing 1950) aufgebaut. Dieser Test dient dazu, allein anhand des Outputs zu prüfen, ob ein Computer zu den gleichen kognitiven Leistungen fähig ist wie ein Mensch. Der Test gilt für den Computer als bestanden, wenn das Ergebnis nicht mehr von dem eines Menschen unterscheidbar ist. Wenn sich nun ein Computer in dem Chinesisch-Zimmer befindet und diesen Test durchläuft, kann geprüft werden, ob er auf chinesische Schriftzeichen, die er als Input erhält, auf die Weise mit chinesischen Schriftzeichen antwortet, die nicht unterscheidbar von der eines Chinesisch sprechenden Menschen ist. Im Fall des Chinesisch-Zimmers, egal ob sich ein Mensch oder Computer darin befindet, würde dieser Test als anstandslos bestanden gelten. Dieser Test ist nach Searle aber nicht hinreichend, um kognitive Leistungen, wie eben das Verstehen der Symbole, feststellen zu können.

Die Person in dem Zimmer manipuliert physische Objekte, wie es auch ein *Physical Symbol System* im Sinne von Newell & Simon nach rein formalen Gesichtspunkten tun würde. Während die Person (oder auch das gesamte Zimmer) nach den Regeln der Anleitung die Symbole manipuliert, weiß sie (oder das ganze Zimmer), so die Sicht, die Searle nahelegen möchte, jedoch nicht, was die Symbole bedeuten. Die Semantik könne also nicht erschöpfend durch eine alleinige Berufung auf Symbole und ihre regelgeleitete Verarbeitung erklärt werden. Daher sei auch eine künstliche Intelligenz im stärksten Sinne nicht möglich, weil hier ein wesentlicher Unterschied zwischen der rein syntaktischen Verfahrensweise und dem menschlichen Verstehen besteht. Vielmehr fehlt der künstlichen Intelligenz sogar die Fähigkeit des Verstehens und bleibt damit hinter den menschlichen Fähigkeiten zurück. Wenn man davon ausgeht, dass Denken als Komputation und regelgeleitete Symbolmanipulation abläuft, dann kann die klassische Kognitionswissenschaft jedoch auch keine gute Erklärung dafür bieten, wie Menschen z.B. Sprache verstehen, da Sprache Bedeutung trage, die sich nämlich nicht gänzlich in den Beziehungen von Symbolen erschöpft. Dieser Mangel klassischer Erklärungen wirft für Harnad folgende Fragen hinsichtlich einer postulierten Symbolverarbeitung auf:

How can the semantic interpretation of a formal symbol system be made intrinsic to the system, rather than just parasitic on the meanings in our heads? How can the meanings of the meaningless symbol tokens, manipulated solely on the basis of their (arbitrary) shapes, be grounded in anything but other meaningless symbols? (Harnad 1990: 335)

Harnad entwickelt eine Erweiterung des Arguments von Searle, das *Merry-Go-Round*-Argument, um schließlich zu zeigen, dass diese Fragen ohne die Annahme einer repräsentationalen Basis nicht zu beantworten sind: Angenommen, jemand soll Chinesisch als Zweitsprache lernen und hat dabei nur ein Chinesisch/Chinesisch-Wörterbuch zur Verfügung. Das Verfolgen der Verweise von einem bedeutungslosen Symbol (oder einer Symbolkette) zum nächsten bedeutungslosen Symbol würde nie zu einem Ende kommen, da man auf diese Weise nicht verstehen würde, was die Symbole bedeuten. Dieses erfolglose Vorgehen gleiche aber währenddessen der Arbeitsweise eines rein symbolbasierten und komputational arbeitenden Systems, das auch keinen Ausgangspunkt oder Basis für die Bedeutung der Symbole finden kann, sondern nur Verweisen von Symbolen auf andere Symbole folgt. Dass nicht klar ist, worin diese Basis besteht und auf welche Weise sie den Symbolen semantischen Gehalt verleihen könnte, ist von Harnad als das Symbol Grounding-Problem formuliert worden:

This is more like the actual task faced by a purely symbolic model of the mind: How can you ever get off the symbol/symbol merry-go-round? How is symbol meaning to be grounded in something other than just more meaningless symbols? This is the symbol grounding problem. (Harnad 1990: 340)

Klassischerweise wird als Lösung dieses Problems geltend gemacht, dass die Bedeutung der Symbole durch ihre angemessene Verbindung zur Welt zustande kommt, durch die Referenz der Repräsentationen. Die Referenz würde dadurch gewährleistet, dass Kognition als ein autonomes funktionales Modul mit peripheren Elementen wie der Sensorik verbunden ist, um die Welt der Objekte zu erkennen, auf die die Symbole verweisen. Harnad sieht hierin aber eine Trivialisierung des Problems, da alle Schwierigkeiten übergangen würden, die mit der Frage zusammenhängen, wie die Symbole sich auf die Dinge und Tatsachen in Welt beziehen können und als Bedeutungsträger fungieren.

Weiter zieht er den Konnektionismus in Erwägung, um zu erklären, wie das Erlernen invarianter Eigenschaften von kategorialen Repräsentationen der Außenwelt möglich ist. Aber auch dieser Ansatz sei nicht ausreichend, da hier allgemein keine Symbole angenommen werden und so auch nicht die Systematizität und Kompositionalität hinsichtlich der Semantik von Gedanken erklärt werden können. Wie bereits Fodor und Pylyshyn am Konnektionismus kritisierten, müssten Symbole angenommen werden, um auch unsere höherstufigen Fähigkeiten wie formale Schlussfolgerungen erklären zu können. Ohne auf die Annahme von Symbolen verzichten zu wollen, schlägt Harnad ein hybrides Modell des Geistes vor:

What will be proposed is a hybrid nonsymbolic/symbolic system, a "dedicated" one, in which the elementary symbols are grounded in two kinds of nonsymbolic representations that pick out, from their proximal sensory projections, the distal object categories to which the elementary symbols refer. (Harnad 1990: 340)

Über die Verbindungen eines neuronalen Netzes hinaus müsse man ein weiteres autonomes symbolisches Modul annehmen, das die kategoriale Zuordnung der einzelnen sensorischen Repräsentationen übernimmt. Nur so sei es möglich, dass auf unterster Ebene einzelne Vorkommnisse von sensorischen Repräsentationen durch kategorialen Symbole höherer Ordnung reguliert werden. Die Symbolmanipulation liefere nicht nur auf Grundlage der einzelnen Vorkommnisse von Repräsentationen ab, sondern auch mittels kategorialer Symbole, die mit den elementaren Symbolen in der Weise verbunden sind, dass sie aus diesen bestehen. Mithilfe dieses hybriden Modells möchte Harnad weiter einen Entwurf für eine mögliche Lösung des Symbol Grounding-Problems anbieten, um zu erklären, wie eine Basis für die symbolischen Repräsentationen in den nicht-symbolischen Repräsentationen zu verstehen sein könnte:

Symbolic representations must be grounded bottom-up in nonsymbolic representations of two kinds: (1) "iconic representations", which are analogs of the proximal sensory projections of distal objects and events, and (2) "categorical representations", which are learned and innate feature-detectors that pick out the invariant features of object and event categories from their sensory projections. Elementary symbols are the names of these object and event categories, assigned on the basis of their (nonsymbolic) categorical representations. Higher-order (3) "symbolic representations", grounded in these elementary symbols, consist of symbol strings describing category membership relations (e.g., "An X is a Y that is Z"). (Harnad 1990: 335)

Menschen besitzen z.B. Fähigkeiten, Objekte zu diskriminieren, zu identifizieren und zu beschreiben, um sich in der Welt zurechtzufinden. Um solche Fähigkeiten zu erklären, nimmt Harnad drei verschiedene Formen von Repräsentationen an: Die ikonischen Repräsentationen (1) sind dabei als analoge Kopien der sensorischen Projektion der Objekte in der Welt zu verstehen, wie beispielsweise das visuelle Bild eines Objektes. Die Ikone dienen der Unterscheidung verschiedener Sachverhalte hinsichtlich ihrer (visuellen/sensorischen) Ähnlichkeiten und Unterschiede. Die Diskriminierung ist aber unabhängig von der Identifikation, da man Objekte diskriminieren kann, ohne sie dabei identifizieren zu müssen. Die ikonischen Repräsentationen wiederum reichen aber nicht aus, um unsere Fähigkeit der Identifikation zu erklären, da der Übergang von Objekten fließend sein könnte. Erst durch eine Rückführung auf bestimmte Invarianten können Objekte einer Kategorie zugeordnet werden und von Objekten anderer Kategorien unterschieden werden. Die kategorialen Repräsentationen (2) seien ikonische Repräsentationen die hinsichtlich der Invarianten bereits gefiltert sind und währenddessen aber die wesentlichen Eigenschaften der sensorischen Projektion bewahren, um so verlässlich für die kategoriale Zuordnung zu sorgen. Sowohl die ikonischen wie auch die

kategorialen Repräsentationen sind dabei noch als rein sensorisch und nicht-symbolisch zu verstehen.

Da beide, (1) und (2), auf sensorischer Ebene arbeiten, gäbe es währenddessen auch keine Probleme hinsichtlich der Referenz auf die Objekte in der Welt. Die referentielle Beziehung der Repräsentationen zu den Objekten sei dabei rein kausal und basiert auf der Relation zwischen distalen Objekten, proximalen sensorischen Projektionen und den erworbenen internen Veränderungen aus der Geschichte von vergangenen Interaktionen mit den Objekten. Währenddessen gäbe es auch weiter keine Probleme mit der Semantik der Repräsentationen, die sich in abgeleiteter Form daraus ergibt, dass sie für etwas eintreten:

Nor is there any problem of semantic interpretation, or whether the semantic interpretation is justified. Iconic representations no more "mean" the objects of which they are the projections than the image in a camera does. Both icons and camera-images can of course be interpreted as meaning or standing for something, but the interpretation would clearly be derivative rather than intrinsic. (Harnad 1990: 343)

Es wird zusätzlich aber noch eine symbolische Ebene benötigt, denn um die kategorischen Repräsentationen manipulieren zu können, müssen diese weiter in ein System übertragen werden, in dem auch neue Bedeutungen durch Kombination der symbolischen Repräsentationen entstehen können. Anders als in einem rein symbolischen System gingen die symbolischen Repräsentationen dabei aber auf die ikonischen und kategorialen Repräsentationen zurück und seien nach Harnad auf diese Weise verankert oder „grounded“.

Zur Veranschaulichung dieser Idee liefert Harnad ein Beispiel: Nehmen wir an, dass die Wörter „Pferd“ und „Streifen“ bereits durch Erfahrung in den ikonischen und kategorischen Repräsentationen verankert sind. Daraufhin könnte weiter auch die Repräsentation eines Zebras aus den Symbolen „Pferd“ und „Streifen“ generiert werden. Die Repräsentation „Zebra“ würde währenddessen die Verankerung durch die anderen beiden symbolischen Repräsentationen erben. Da „Pferd“ und „Streifen“ ikonisch und kategorisch bereits verankert sind, sei nach Harnad ebenso das neu erzeugte Symbol „Zebra“ verankert. Jemand, der lediglich die Bedeutung der Wörter „Pferd“ und „Streifen“ kennt, sei dann auch in der Lage, das neue Symbol „Zebra“ bei der ersten Begegnung mit einem Objekt in der Welt zu identifizieren. Sobald man über eine sensorisch verankerte Menge an elementaren Symbolen verfügt, können die weiteren Symbolketten einer natürlichen Sprache allein durch kompositionale Zusammenstellungen generiert werden.

The expectation has often been voiced that "top-down" (symbolic) approaches to modeling cognition will somehow meet "bottom-up" (sensory) approaches somewhere in between. If the grounding considerations in this paper are valid, then this expectation is hopelessly modular and there is really only one viable route from sense to symbols: from the ground up. (Harnad 1990: 345)

Gegenüber klassischen Ansätzen, bei denen wie im Fall der Hypothese eines physischen Symbolsystems oder der LoTH eine top-down-Erklärung vorgenommen wird und man direkt von der Symbolmanipulation ausgeht, möchte Harnad einen bottom-up-Ansatz etablieren, bei der explanatorisch mit den sensorischen Bedingungen begonnen wird. Es wird hier also ein Bild skizziert, bei dem die sensorische Basis als Bedingung der Symbolverarbeitung hervorgehoben wird. Dieses Bild steht beispielsweise Fodors Erklärung entgegen, da hier bei der Beschreibung bei den Repräsentationen angesetzt wird, die bereits eine bestimmte Übersetzung erfahren haben und nicht mehr im direkten Zusammenhang mit der Sensorik gesehen werden.



<http://www.springer.com/978-3-658-17218-3>

Die körperliche Konstitution von Kognition

Weber, A.M.

2017, X, 290 S. 5 Abb., Softcover

ISBN: 978-3-658-17218-3