Metzinger · Subjekt und Selbstmodell

Thomas Metzinger

Subjekt und Selbstmodell

Die Perspektivität phänomenalen Bewußtseins vor dem Hintergrund einer naturalistischen Theorie mentaler Repräsentation

2., durchgesehene Auflage

mentis
Paderborn

Umschlaggestaltung unter Verwendung eines Bildes von:
Jean Metzinger, Die Bucht von Saint-Tropez (ca. 1904).

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Metzinger, Thomas:
Subjekt und Selbstmodell: die Perspektivität phänomenalen
Bewußtseins vor dem Hintergrund einer naturalistischen Theorie
mentaler Repräsentation / Thomas Metzinger. – 2., durchges. Aufl. –
Paderborn: mentis, 1999
ISBN 3-89785-081-8

Einbandgestaltung: Graphik-Studio Winkler, Stuttgart

Gedruckt auf umweltfreundlichem, chlorfrei gebleichtem
und alterungsbeständigem Papier ♾ ISO 9706

© 1999 mentis Verlag, Paderborn
(mentis Verlag GmbH, Schulze-Delitzsch-Straße 19, D 33100 Paderborn)

Alle Rechte vorbehalten. Dieses Werk sowie einzelne Teile desselben sind urheberrechtlich geschützt. Jede Verwertung in anderen als den gesetzlich zugelassenen Fällen ist ohne vorherige schriftliche Zustimmung des Verlages nicht zulässig.

Printed in Germany
Satz: Graphische Werkstätten Lehne GmbH, Grevenbroich
Druck: WB-Druck, Rieden/Allgäu
ISBN 3-89785-081-8

Für meine Eltern

Inhalt

Vorwort 9

1. Kapitel 13
Auf dem Weg zu einer neuen Theorie des Geistes 13

1.1 Einleitung 13
1.2 Subjektivität als Kernstück einer modernen Theorie des Geistes: Die drei Hauptelemente des Problems 21
1.3 Kriterien für eine Theorie der Subjektivität: *Die Fragen* ... 38

2. Kapitel
Mentale Repräsentation und phänomenale Zustände 47

2.1.1 Mentale Repräsentation: Intentionalität durch Informationsverarbeitung 47
2.1.2 Mentale Simulation: Die Erzeugung virtueller Erlebniswelten durch Phantasie, Imagination, Planung 65
2.1.3 Mentale Präsentation: Qualia 72
2.1.4 Subsymbolische Meta-Repräsentation: Bewußtsein 87
2.2 Mentale Modelle: Die abstrakten Werkzeuge biologischer Systeme 104
2.2.1 Digitales Wissen und analoges Wissen 125
2.3 Gehirne als Generatoren von Modellen der Wirklichkeit ... 135
2.3.1 Repräsentationale Gesamtzustände 135
2.3.2 Deviante mentale Modelle der Welt: Agnosien, Neglekte, Blindsicht, Halluzinationen und Träume 137
2.3.3 Monozentrierte Weltmodelle 149

3. Kapitel
Die Selbstmodell-Theorie der Subjektivität 151

3.1.1 Mentale Selbstrepräsentation: Mereologische Intentionalität 151
3.1.2 Mentale Selbstsimulation: Interne Selbstähnlichkeit und Entwürfe des Selbst 155
3.1.3 Mentale Selbstpräsentation: Verkörperung 156
3.2 Was ist ein mentales Selbstmodell? 158
3.2.1 Selbstsimulation und Selbstreferenz 169
3.2.2 Deviante mentale Modelle des Selbst: Ich-Störungen, halluzinierte Selbste, multiple Selbste und luzide Träume 177
3.3 „SMT“: Die Selbstmodell-Theorie der Subjektivität 204

4. Kapitel
Fledermäuse, objektive Selbste und die Irreduzibilität der Innenperspektive 210

4.1 *Inneres Wissen:* Thomas Nagels Philosophie der Subjektivität 210
4.2.1 *„Wie es ist, ein X zu sein“* 211
4.2.2 Perspektivische Fakten 221
4.3 Schwierigkeiten der Nagelschen Subjektphilosophie 230

5. Kapitel
Vom Subjekt zum Selbstmodell: Perspektivität ohne Ego 241

5.1 Der totale Flugsimulator und der kleine rote Pfeil 241
5.2 Subjektivität in zentrierten Repräsentationsräumen: *Vorläufige Antworten* 245
5.3 Mentale Selbstmodellierung und die Würde informationsverarbeitender Systeme 284

6. Kapitel
Anhänge 292

6.1 Literaturverzeichnis 292
6.2 Personenverzeichnis 308
6.3 Begriffsverzeichnis 311

„Der Spaß ist, wenn mit seinem eignen Pulver der Feuerwerker auffliegt . . .“

Shakespeare, *„Hamlet“*

Vorwort

In dieser Arbeit geht es nicht um eine weitere modische Liquidation des Subjekts, sondern um dessen Rehabilitation als Gegenstand ernsthafter theoretischer Bemühungen. Das Thema, um das die folgenden Untersuchungen kreisen, ist die *Perspektivität* unseres phänomenalen Bewußtseins. „Perspektivität“ ist zunächst jedoch nicht mehr als eine metaphorische Anleihe aus der Phänomenologie des visuellen Sinnes: Unser visuelles Erleben der Umwelt ist um ein Zentrum herum aufgebaut, denn als Sehende erleben wir die Welt scheinbar von einem Standpunkt aus. Dieser Standpunkt, der *Stationis Punctum*, scheint hinter unseren Augen zu liegen und er ist der Mittelpunkt unseres visuellen Erlebnisraumes. Von ihm können wir imaginäre Linien zu den Grenzen der Dinge ziehen, die wir in den Blick nehmen. Parallele Geraden scheinen dabei in der Ferne zusammenzutreffen und weiter entfernte Gegenstände sind kleiner als näherliegende. Dieses strukturelle Merkmal des durch unseren Sehsinn erzeugten Bildes der Welt – die Tatsache also, daß es um ein *Zentrum* herum aufgebaut ist – vermittelt uns manchmal den Eindruck, als gäbe es einen kleinen Homunkulus, der hinter unseren Augen sitzt und durch sie wie durch zwei Fenster in die Welt hinausblickt. Wir wissen natürlich, daß dieser Homunkulus hinter den Fenstern mit Blick in die Welt eine Fiktion ist. Aber trotzdem können wir uns nicht von der durch unser Raumerleben erzeugten Illusion befreien, daß unser *Selbst* eine räumliche Lokalisierung besitzt – als Mittelpunkt der visuellen Welt.

Was für den strukturellen Aufbau des Gesichtsfeldes und des visuellen Modells der Wirklichkeit gilt, trifft in einem umfassenderen Sinn auch auf unseren phänomenalen Raum als Ganzen zu: Unser gesamtes Bewußtsein ist um einen Mittelpunkt herum aufgebaut, weil es ein *zentriertes* Bewußtsein ist. Dieser unhintergehbare Mittelpunkt unseres inneren Erlebnisraums ist das phänomenale Selbst, das Subjekt psychischer Zustände. Das Subjekt psychischer Zustände bildet den thematischen Kern der nun folgenden Überlegungen, weil es verantwortlich ist für die „Perspektivität“ unseres mentalen Innenlebens. Unsere innere Biographie ist perspektivisch, weil sie fast immer an ein Erlebnissubjekt gebunden ist, dessen Biographie sie ist. Die Struktur phänomenalen Bewußtseins besser zu verstehen und zu untersuchen, ob Selbstbewußtsein – so wie wir alle es *erleben* – einer naturalistischen Erklärungsstrategie zugänglich sein könnte, ist das zentrale Anliegen meiner Untersuchung.

Die gegenwärtige Diskussion in der analytischen Philosophie des Geistes ist weitverzweigt und teilweise hochspezialisiert. Besonders im angelsächsischen Raum zeichnet sie sich nicht nur durch eine große argumentative Dichte und ein steigendes Tempo aus, sondern auch durch reichhaltige und für den Laien schwer zu überschauende interdisziplinäre Verflechtungen.

Aus diesem Grund habe ich dem interessierten Leser im Fußnotenapparat Hinweise auf wichtige Literatur, deutsche Übersetzungen, weitere zentrale Diskussionsstränge und auf dem Laien leicht zugängliches empirisches Material gegeben. Diese knappen Angaben sollen ein Service für all diejenigen Leser sein, die Schwierigkeiten mit der in dieser Arbeit zum Teil vorausgesetzten Terminologie haben, spezielle Interessen verfolgen oder ihre Kenntnisse weiter vertiefen möchten.

Im ersten Kapitel geht es nach einer sehr kurzen Einführung in die Grundprobleme, mit denen sich eine Philosophie des Geistes gegenwärtig auseinandersetzen muß, um eine klärende Sondierung des Problemfeldes und eine Formulierung der Fragen, die für eine heutige Theorie des Subjekts relevant erscheinen. Der am Ende des ersten Kapitels aufgestellte Fragenkatalog dient als roter Faden, der am Ende wieder aufgenommen wird. Das zweite Kapitel untersucht, wie einzelne Elemente einer empirisch plausiblen Theorie der mentalen Repräsentation gegenwärtig aussehen könnten und ob diese Elemente mögliche Bausteine für eine *philosophische* Theorie des Geistes bilden. Um besser zu verstehen, welche neuen Perspektiven die Fortschritte der Neuro- und Kognitionswissenschaften auf klassische Probleme eröffnen, war es nicht nur wichtig, einige Fallbeispiele aus der empirischen Psychologie näher zu betrachten, sondern auch die logische Struktur der Repräsentationsbeziehung und einige erkenntnistheoretische Fragen in die Überlegungen mit einzubeziehen. Im dritten Kapitel formuliere ich die zentrale These der Arbeit, nämlich die „Selbstmodell-Theorie der Subjektivität". Dabei kommen Einsichten aus dem vorangegangenen Kapitel zur Anwendung auf den Sonderfall der mentalen *Selbstrepräsentation*. Auch hier erschien mir neben der philosophischen Begriffsklärung ein Blick auf neuere empirische Erkenntnisse bezüglich der neurobiologischen Rahmenbedingungen verschiedener Formen von phänomenalem Selbstbewußtsein unerläßlich. Im vierten Kapitel entsteht dann aus einer kurzen Analyse wichtiger antinaturalistischer Argumentationslinien innerhalb der aktuellen philosophischen Diskussion ein zweiter roter Faden in Gestalt einer zweiten Gruppe von Fragen an eine moderne Theorie des Erlebnissubjekts. Um zusätzliche Perspektiven auf die theoretische Problematik zu gewinnen, habe ich dort als „zweites Sprungbrett" in die Diskussion die Thesen eines prominenten Philosophen der angelsächsischen Tradition eingesetzt, weil sie in prägnanter und nichttechnischer Weise zentrale Aspekte der Problemlandschaft beleuchten. Das Schlußkapitel schließlich verknüpft die beiden roten Fäden wieder miteinander und gibt aus der Perspektive der „Selbstmodell-Theorie der Subjektivität" einige vorläufige Antworten auf die Fragen des eingangs entwickelten Katalogs. Hier kehren die Überlegungen sowohl zu den eben bereits angesprochenen Thesen Thomas Nagels bezüglich der Irreduzibilität des subjektiven Standpunkts, die ich im vierten Kapitel zu einem tieferen Eindringen in die Problematik benutze, als auch zu den allgemein-systematischen Fragen des ersten Kapitels zurück. Die Einleitung soll eine vorsichtige Hinführung zu diesem ersten Blick auf die Problemlandschaft sein.

Aus der Vielzahl der Personen, denen ich für ihre Hilfe und Unterstützung bei der Durchführung dieser Arbeit zu Dank verpflichtet bin, möchte ich nur drei hervorheben. Prof. Dr. Werner Becker hat einen großen Anteil am Zustandekommen der vorliegenden Arbeit. Er hat mich beraten und die Rahmenbedingungen geschaffen, unter denen die mehrjährigen Vorarbeiten zu diesem Buch möglich wurden. Mein wichtigster Kritiker und Diskussionspartner in dieser Zeit war Dr. Andreas Bartels. Von ihm habe ich viel gelernt. Er hat häufig einen entscheidenden Einfluß auf den Verlauf meiner Überlegungen gehabt – für ihre Mängel ist er nicht verantwortlich. Und schließlich bin ich wieder einmal Prof. Dr. Peter Bieri für seine kompetente Kritik zu Dank verpflichtet.

Gießen, April 1993 Thomas Metzinger

Vorwort zur 2. Auflage

Einige Aspekte des nun wieder vorliegenden Textes erscheinen mir heute in theoretischer Perspektive und bereits aus der kurzen zeitlichen Distanz zur Erstauflage als unvollständig oder unbefriedigend. Insbesondere die Entwicklung des Erkenntnisfortschritts in den eher empirischen Bereichen der Bewußtseinsforschung ist seit dem ersten Erscheinen von „Subjekt und Selbstmodell“ auf eine sehr ermutigende – und die Hintergrundannahmen der Selbstmodell-Theorie deutlich stützende – Weise, aber auch in einer inhaltlich kaum noch zu überschauenden Dichte und Komplexität verlaufen. Trotzdem habe ich mich nach längerem Zögern und wegen der anhaltenden Nachfrage nach „Subjekt und Selbstmodell“ entschlossen, einer Neuauflage zuzustimmen. Die vollständige philosophische Umsetzung der vielen neuen Gesichtspunkte, die sich sowohl auf begrifflicher Ebene als auch in empirischer Hinsicht seit dem ersten Erscheinen des Buchs ergeben haben, wird jedoch einer zukünftigen Fassung der Theorie vorbehalten bleiben müssen.

San Diego, im Mai 1999

1. Kapitel

Auf dem Weg zu einer neuen Theorie des Geistes

1.1 Einleitung

Wir befinden uns auf dem Weg zu einer neuen Theorie des Geistes. Seit geraumer Zeit bewegen wir uns mit steigender Geschwindigkeit auf ein neues theoretisches Verständnis unserer inneren Natur zu. Dieses neue theoretische Verständnis des Geistes wird nicht das erste seiner Art sein, vielmehr stellt es eine weitere Station auf einem langen und verschlungenen Entwicklungsweg intellektueller Traditionen dar. Die heutige Situation ist jedoch gegenüber früheren Stationen der Problem- und Begriffsgeschichte grundlegend verändert. Wir sind mit einer in der Geschichte der Menschheit bisher einmaligen Situation konfrontiert: Noch nie gab es ein so umfangreiches und auch für den Geisteswissenschaftler leicht verfügbares empirisches Wissen über die objektiven Entstehungsbedingungen und Eigenschaften subjektiver Zustände. Das Wissen über die physischen Rahmenbedingungen psychischer Zustände wächst zudem mit steigender Geschwindigkeit an und es ist abzusehen, daß dieser Wissenszuwachs im kommenden Jahrhundert einen dramatischen Charakter annehmen wird. Schon jetzt ist eine immer größer werdende Anzahl häufig neu entstehender Einzeldisziplinen – wie etwa die Psychobiologie, die Neuroinformatik oder die Künstliche-Intelligenz-Forschung – an einer nur schwer zu überschauenden interdisziplinären Diskussion beteiligt. In dieser Diskussion steht die Philosophie häufig etwas isoliert da und kämpft gegen nicht immer unberechtigte Vorurteile seitens empirischer Wissenschaftler, die ihre Seriosität oder die Brauchbarkeit ihrer Beiträge für den Fortschritt des Projekts überhaupt betreffen. Auf der anderen Seite werden Struktur und Zielrichtung der Diskussion in den Neuro- und Kognitionswissenschaften für all jene immer unübersichtlicher, die sich eine allgemeine Einordnung der neuen Erkenntnisse und Einsichten über ihr eigenes subjektives Innenleben erhoffen. Auf diese Weise ist auch das traditionelle Beziehungsgefüge zwischen Geistes- und Naturwissenschaften ins Wanken geraten.

In der Geschichte der Menschheit sind eine große Anzahl von Theorien des Geistes entwickelt worden. Diese Theorien waren mythologische, theologische oder philosophische Versuche, einen Aspekt des Menschen und auch der Welt überhaupt zu erklären, der alle bloß äußeren Merkmale zu überschreiten scheint. Denn es scheint etwas zu geben, das hinter den Dingen liegt, ihre Erkenntnis erst möglich macht und vielleicht sogar vor ihnen immer schon da war. Nach traditionellem Verständnis kommt dieses verborgene und doch in gewissem Sinne allem zugrundeliegende Etwas

gerade menschlichen Wesen in einer ganz besonderen Weise zu: Wir begreifen uns selbst als *geistige* Wesen.

Theorien des Geistes hat es nicht immer gegeben. In der abendländischen Geistesgeschichte gab es so etwas wie einen Zeitpunkt der Entdeckung des Geistes. Mit der Entdeckung des *nous* in der Antike fand der Geist seinen Eingang in die Diskussionen der Philosophen. Seitdem hat er sie als Phänomen und als Problem immer wieder beschäftigt, er wurde zu einem beständigen *topos* des philosophischen Denkens im Okzident. In unserem Kulturkreis ist die Geschichte des Geistbegriffs die Geschichte einer Entwicklung von mythisch-sinnlich-konkreten Bedeutungsinhalten über eine personal-christliche Interpretation bis hin zu einem abstrakt-spekulativ-metaphysischen Sinn, der schließlich im Hegelschen System gipfelt. Es ist die Geschichte einer immer weiter fortschreitenden semantischen Differenzierung, in deren Verlauf der Begriff des Geistes zu einem der grundlegendsten philosophischen Konzepte überhaupt wurde.

Heute befinden wir uns erneut auf dem Weg zu einer umfassenden Theorie des Geistes. Obwohl jedoch durch die beeindruckende Zunahme unseres empirischen Wissens bezüglich der physischen Entstehungsbedingungen von Bewußtsein und Subjektivität zunächst berechtigte Hoffnungen auf klar benennbare Erkenntnisfortschritte in der nahen Zukunft geweckt werden, scheinen auf der anderen Seite die Komplexität, das Tempo und die Unübersichtlichkeit der Debatte in den Einzelwissenschaften diese Hoffnungen wieder zunichte zu machen. Außerdem breiten sich Befürchtungen aus, im Rahmen der geschilderten Entwicklung könnte es zu einer szientistischen Kolonisierung unserer Innenwelt kommen, in deren Verlauf unser lebensweltliches Selbstverständnis auf kränkende Weise *ad absurdum* geführt wird und an deren Ende der Reichtum und die Vielfalt unseres Innenlebens durch die Metaphern eines neuen materialistischen Jargons einplaniert werden. Trotzdem beginnt sich das neue Bild des Geistes bereits in ersten Umrissen abzuzeichnen und es wird auch deutlich, daß dieses Bild fundamentale Konsequenzen für unsere Anthropologie haben wird.

Was ist die Rolle der Philosophie in einer solchen Situation? Philosophie des Geistes ist heute unter anderem ein *metatheoretisches* Projekt. Das heißt: Philosophie ist gegenwärtig häufig ein theoretisches Unternehmen zweiter Ordnung, eine Theorie über andere Theorien. Philosophie des Geistes zu betreiben bedeutet, an einer Metatheorie der empirischen Psychologie, der Neuro- und Kognitionswissenschaften zu arbeiten. Diese philosophische Metatheorie sollte mindestens zweierlei leisten. **Erstens** erwarten wir uns von ihr eine Zusammenschau: Sie sollte die Fülle der vorläufigen Erkenntnisse ihrer Zeit in ein allgemeineres Bild fassen können. Da empirische Erkenntnisse sich in einem geschichtlichen Fluß befinden, wird dieses Bild immer nur ein provisorisches Bild sein, eine vorläufige Darstellung, die jederzeit durch neue Entdeckungen revidiert werden kann. Dieses vorläufige synoptische Bild des Phänomens „Geist“ soll uns helfen bei der Beantwortung genereller philosophischer Fragen wie: Was ist das Wesen

von Erkenntnis? Was bedeutet es, daß der Mensch ein geistiges Wesen ist? Wie haben wir uns die Beziehung zwischen körperlichen und geistigen Vorgängen zu denken? Gibt es einen unveränderlichen Wesenskern im Menschen? **Zweitens** besteht die gegenwärtige Aufgabe der Philosophie des Geistes darin, eine begrifflichen Kommentar zu den Entwicklungen der einzelwissenschaftlichen Theorien anzubieten. Dieser Kommentar muß die durch den empirischen Wissenszuwachs ausgelösten Veränderungen in unserem Weltbild widerspiegeln. Er tut dies zum Beispiel, indem er auf notwendig gewordene Revisionen unseres Geistbegriffs hinweist oder überflüssig gewordene Konzepte eliminiert und neue entwickelt.

Das Projekt einer Theorie des Geistes besitzt einen Doppelcharakter. Das ist deshalb so, weil es sich in unserer Zeit notgedrungen aus empirischen und metatheoretischen Anstrengungen zusammensetzt. Die Rolle der empirischen Forschung besteht darin, ein genaueres Verständnis der Rahmenbedingungen für das Entstehen psychischer Zustände zu erarbeiten: Sie muß die *Mechanismen* untersuchen, die der Entstehung von Bewußtsein zugrunde liegen. Ihr Ziel ist auf der einen Seite die Erarbeitung von *kausalen* Erklärungen, also von Erkenntnissen über Gesetzmäßigkeiten oder auch darüber, warum ein bestimmtes System gewisse psychologische Eigenschaften instantiiert. Die Rolle des metatheoretischen Kommentars auf der anderen Seite besteht darin, eine begleitende Begriffsanalyse durchzuführen und die im Entstehen begriffene Theorie des Geistes in ihren jeweiligen Entwicklungsstadien auf innere Konsistenz zu überprüfen. Denn je besser wir unsere subjektive, innere Wirklichkeit und ihre Wechselwirkungen mit der objektiven, äußeren Wirklichkeit zu verstehen beginnen, um so engmaschiger und subtiler wird auch das Begriffsnetz sein müssen, das wir über diese beiden Wirklichkeiten werfen – um sie doch niemals vollständig einzufangen. Die Anpassung dieses Netzes aus Begriffen an eine rapide expandierende Datenmenge benötigt darum die begriffswissenschaftliche Begleitung einer Philosophie des Geistes, die analytische Professionalität mit einem ernsthaften Erkenntnisinteresse und ideologischer Ungebundenheit verbindet.

Allerdings dürfen philosophische Diskurse eines solchen interdisziplinären Typs nicht parasitär an empirische Debatten angelagert werden und lediglich als Sprungbrett für ein analytisches *l'art pour l'art* dienen. Bisher waren es überwiegend analytische Philosophen des Geistes aus der angelsächsischen Tradition, die die wesentlichen Beiträge zum Aufbau einer gehaltvollen, postmetaphysischen Theorie des Geistes geliefert und die Diskussion durch ihre Dominanz geprägt haben. Und in manchen Fällen sind sie auch der oben angedeuteten Versuchung erlegen. Betrachtet man jedoch die theoretische Entwicklung seit dem Zweiten Weltkrieg in diesem Bereich insgesamt, dann kann man die Fortschritte gegenüber früheren Stationen der Problemgeschichte kaum übersehen: Die Fragestellungen haben durch die deutlich wahrzunehmende Präzisierung eine starke Differenzierung erfahren. Das Niveau von Analysen und Antwortversuchen hat sich in der Folge stetig angehoben.

Als Philosophen sind wir aber derzeit weit entfernt davon, Antworten auf die vielen Probleme im Kontext einer neuzeitlichen Theorie des Geistes liefern zu können. Im Gegenteil – die präzise Analyse der *Fragestellungen* muß zum gegenwärtigen Zeitpunkt in unseren Überlegungen eine mindestens ebenso große Rolle spielen wie die empirische Suche nach möglichen Teilantworten und Lösungselementen. In einer solchen Situation sind zwei zentrale Kriterien maßgeblich für die Beurteilung provisorischer Versuche. **Erstens** muß man fordern, daß solche Versuche *nicht mit unserem derzeitigen empirischen Wissen kollidieren.* Eine Theorie über mentale Zustände darf nicht in Widerspruch stehen zum jeweiligen Faktenwissen über die physischen Hintergrundbedingungen solcher Zustände. Sie sollte empirisch plausibel sein und das ganze Spektrum subjektiver Bewußtseinszustände in ihrem gesamten phänomenologischen Reichtum verständlich machen können. Und **zweitens** muß der dieses Faktenwissen synthetisierende begriffliche Kommentar *konsistent* sein. Eine – wie auch immer vorläufige – Theorie des Geistes muß logischer Überprüfung standhalten können und den größtmöglichen Teil unseres empirischen Wissens erklärend integrieren.

Wie könnte eine moderne Theorie des Geistes aussehen? Die zukünftige Theorie wird eine Theorie über mentale Zustände sein: Sie wird uns das Wesen von Wünschen, Meinungen, Gefühlen oder Wahrnehmungserlebnissen erklären. Dabei interessiert an erster Stelle, von was mentale Zustände Zustände sind – sind mentale Zustände die Zustände eines nichtphysischen Einzeldings, etwa die Zustände einer Seele – oder sind sie bloß Zustände unserer Gehirne, vielleicht die Zustände einer von ihnen aufgebauten Datenstruktur? Wenn das Projekt einer neuen Theorie des Geistes gelingen soll, so muß uns außerdem in seinem Rahmen verständlich werden was es heißt, daß wir Wesen mit einem *Innenleben* sind. Denn wir sind Wesen im Besitz einer psychischen Biographie, welche ihnen eine zugleich rätselhafte und würdevolle Perspektive auf sich selbst und die Welt verleiht. Eine erfolgreiche Theorie des Mentalen müßte auf diese Weise auch ihren Beitrag liefern zu einer philosophischen Anthropologie, die ein überzeugendes Bild des Menschen als eines Innerlichkeit besitzenden Wesens anzubieten hat. Weiterhin muß sie Auskunft geben über die Möglichkeit und das Zustandekommen der durch diese Innerlichkeit mediatisierten Erkenntnisleistungen, über die *kognitive Rolle*, die manche mentalen Prozesse in unserer „epistemischen Ökologie“ spielen, in der Organisation des komplizierten Netzwerks von Wissensbeziehungen, durch das wir in unsere Umwelt eingebunden sind. Sie soll aber gleichzeitig von der spezifisch menschlichen Perspektive abstrahieren und nicht in einem psychologischen Speziezismus gefangen bleiben. Deshalb muß eine Theorie des Geistes, wenn sie befriedigend sein will, auch die Frage nach der Möglichkeit einer *universellen Psychologie* – die auch Tiere, außerirdische oder künstliche Systeme erfassen könnte – und ihrer Beziehung zu einer Erkenntnistheorie im allgemeinen beantworten können. Und schließlich erhoffen wir uns von einem solchen metatheoretischen, die Erkenntnisse der Einzelwis-

senschaften in einem philosophischen Kommentar synthetisierenden Unternehmen auch eine Erläuterung der kausalen Beziehung zwischen mentalen und physischen Phänomenen.

Dabei sind es drei große Problemfelder, an denen sich eine philosophische Theorie des Geistes zu orientieren hat: Die *Intentionalität* des Mentalen, das *Leib-Seele-Problem* und die Frage nach dem *Subjekt.* Werfen wir deshalb einen kurzen Blick auf diese drei wesentlichen Klassen von philosophischen Problemen, denen wir auf dem Weg zu einer Theorie des Geistes begegnen werden.

Die Intentionalität des Mentalen. Mentale Phänomene besitzen einen *Gehalt*, sie sind Zustände, die auf Bereiche der Welt gerichtet sind und sie in einem rätselhaften Sinne enthalten: Mentale Zustände sind Zustände, die von etwas außerhalb ihrer selbst bestimmt oder geprägt werden. Eine der wesentlichsten Einsichten über phänomenales Bewußtsein scheint nämlich zu sein, daß man das Prädikat *„bewußt"* immer auch als zweistelliges Prädikat analysieren kann – *Bewußtsein ist Bewußtsein von etwas.* In der philosophischen Tradition ist diese Eigenschaft als kategoriales Abgrenzungskriterium des Mentalen gegenüber dem Physikalischen diskutiert und entwickelt worden. Was uns als geistige Wesen von bloßen physikalischen Gegenständen oder einfachen Organismen unterscheidet, ist die kaum bezweifelbare Tatsache, daß einige unserer inneren Zustände über sich selbst hinaus verweisen indem sie eine Bedeutung besitzen.

Bei kognitiven mentalen Zuständen mit propositionalem Gehalt wird das philosophische Kernproblem besonders deutlich: Was haben alle Personen miteinander gemein, die glauben, *daß Franz Brentano der Autor eines Werkes mit dem Titel „Psychologie vom empirischen Standpunkt" war*? Sie scheinen sich alle in demselben mentalen Zustand zu befinden. Dieser Zustand ist ein Zustand vom Typ „*Glauben* (oder: meinen, denken, bezweifeln, wünschen usw.), *daß p*". Dabei bezieht *p* sich auf eine mögliche Wahrheit, die in einem Satz einer natürlichen Sprache ausgedrückt werden könnte. Zum Beispiel könnte *p* für die Wahrheit stehen, *daß Franz Brentano der Autor eines Werkes mit dem Titel „Psychologie vom empirischen Standpunkt" war.* Es ist nun allerdings unwahrscheinlich, daß alle Personen, die diese Wahrheit erfassen, ein gemeinsames konkretes Merkmal besitzen – etwa ein bestimmtes physikalisches Merkmal ihrer Gehirnzustände oder einen speziellen kausalen Mechanismus, der ihr äußeres Verhalten steuert. Was aber ist es dann, das alle denkbaren Wesen im Universum – Marsmenschen, Großrechner oder Engel –, deren subjektive Zustände denselben intentionalen Gehalt besitzen, miteinander verbindet? Wie entscheiden wir, ob zwei mentale Zustände *identisch* sind?

Eine befriedigende Theorie des Geistes muß uns aber nicht nur erklären können, was die Identitätskriterien bezüglich zweier mentaler Zustände sind. Sie muß uns auch darüber aufklären, ob es tatsächlich unsere Meinungen und unsere Überzeugungen sind, die unsere Handlungen verursachen. Denn die unser vorphilosophisches Selbstverständnis prägende Alltagspsychologie geht davon aus, daß es der intentionale Gehalt unserer geistigen

Zustände ist, der die *kausale Rolle* bei der Verursachung unseres Verhaltens spielt: Wir nehmen ein bestimmtes Buch noch einmal aus dem Regal, **weil** wir glauben, *daß Franz Brentano der Autor eines Werkes mit dem Titel „Psychologie vom empirischen Standpunkt" war.* Für unsere Analyse anderer Menschen als vernünftiger Wesen spielen solche Annahmen über das Wesen ihrer geistigen Zustände eine wichtige Rolle. Wie aber kann die Intentionalität des Mentalen in kausale Verhaltenserklärungen eingehen, wenn doch anzunehmen ist, daß es keinen scharf umgrenzten Set physischer Eigenschaften gibt, der alle Systeme miteinander verbindet, die sich in bestimmten intentionalen Zuständen befinden?

Die Tatsache, daß mentale Zustände einen intentionalen Gehalt besitzen, führt automatisch zu einer engen Verbindung zwischen der Philosophie des Geistes und der Erkenntnistheorie. Es hat den Anschein, als ob wir mit der Hilfe unserer inneren Zustände Wahrheiten erfassen können. Mentale Zustände fungieren als Träger von Wissen über die Welt. Wenn das so ist, dann muß eine naturalistische Theorie des Geistes – eine Theorie also, die davon ausgeht, daß unser Innenleben letztlich ein natürlich entstandenes Phänomen ist, für das natürliche Erklärungen gefunden werden können – Auskunft darüber geben, wie das mental repräsentierte Wissen über die Welt *in dieser Welt* sein kann: Wie können mentale Zustände – inklusive ihres vollen Gehalts – selbst ein Teil der Welt sein? Als Mediatoren kognitiver Prozesse werden sie damit auch wichtig für ein anderes Projekt zeitgenössischen Philosophierens: Für die Naturalisierung der Erkenntnistheorie durch ihre Neuinterpretation im Rahmen empirischer Kognitionstheorien. Wenn solche Theorien einmal erfolgreich sein sollten, werden sie uns verdeutlichen, wie das Mentale durch epistemische Prozesse – also durch Vorgänge der Wissensgewinnung – in die Welt eingebunden ist. Das **zweite** große Feld philosophischer Fragen ist in seiner expliziten Form sogar noch älter als das Problem der Intentionalität des Mentalen – seine thematische Grundlage bildet die Möglichkeit der Einbindung mentaler Prozesse in die Welt durch Ursachenketten.

Das Leib-Seele-Problem. Weiß ich eigentlich genau, was ich meine, wenn ich sage, daß meine Entscheidung, meinen Arm zu heben, die Bewegung des Arms *hervorruft*? Und wenn ich sage, daß die sinnliche Wahrnehmung eines Gegenstandes in mir gewisse Gedanken und Gefühle auslöst: Folge ich dann nur einer allgemein verbreiteten Sprechweise (nämlich dem christlich-cartesianisch geprägten Alltagsidiom) oder beschreibe ich einen Vorgang, den ich in mir *beobachten* kann, nämlich die Auslösung oder Verursachung physischer Ereignisse durch mentale Ereignisse? Solche Fragen spiegeln den Übergang von der intuitiv-vorphilosophischen in die eigentlich theoretische Dimension des Leib-Seele-Problems wider. Diese Fragen würden sich für Wesen mit Formen des Bewußtseins und Selbsterlebens, die von den unseren radikal unterschieden sind, vielleicht gar nicht oder ganz anders stellen.

Ein ontologischer Dualismus nach dem Modell der cartesianischen Substanzentheorie – er stellt im abendländischen Kulturkreis den pointierte-

sten und gleichsam paradigmatischen Lösungsversuch des Problems dar – hebt die Leib-Seele-Frage bereits auf eine explizit theoretische Ebene. Aus heutiger Perspektive erscheint er jedoch als eine unhaltbare Position; und das aus einer Reihe von Gründen. Einer dieser Gründe ist zum Beispiel das *Prinzip der kausalen Geschlossenheit der physikalischen Welt*: Um ernsthaft an einer Wechselwirkungstheorie zwischen Mentalem und Physischem im Menschen festhalten zu können, müßten wir annehmen, daß im menschlichen Gehirn nicht-physikalische Ursachen und Wirkungen zu beobachten sind. Tatsächlich werden solche Thesen auch heute noch von Philosophen vertreten. Sie kollidieren jedoch mit unseren besten, empirisch fundierten Theorien über die physikalische Welt: Es gibt einfach keinen Grund zu der Annahme, daß ausgerechnet im menschlichen Gehirn der Energieerhaltungssatz verletzt sein sollte.

Wenn man die moderne Diskussion des Leib-Seele-Problems in der Zeit nach dem Zweiten Weltkrieg analysiert, dann stellt man fest, daß eine Fülle neuer Theorien bezüglich des psychophysischen Problemkomplexes entwickelt worden ist. In der neueren Diskussion gibt es etwa neun verschiedene Typen von theoretischen Ansätzen. Es fällt dabei auf, daß die unterschiedlichen philosophischen Positionen zum Leib-Seele-Problem sich heutzutage teilweise auf einem recht hohen analytischen Niveau bewegen. Ins Auge sticht zudem, daß – mit Ausnahme des interaktionistischen Dualismus von Karl Popper und Sir John Eccles und einiger mit ihm verwandter Thesen – der Physikalismus zur orthodoxen Hintergrunddoktrin geworden ist. Trotzdem kann keine der aktuellen materialistischen Theorien als befriedigende Lösung der Probleme gelten. Das gilt auch für die funktionalistischen (also an der Computermetapher orientierten, das heißt: das Verhältnis von Geist und Körper analog zu dem von *Software* und *Hardware* interpretierenden) Versuche, die Essenz des Mentalen begrifflich einzufangen. Eines der Kernprobleme in der heutigen Diskussion drückt sich etwa in der folgenden Frage aus: Wir wissen, daß es deutliche Korrelationen zwischen mentalen Ereignissen (den Zuständen von *Personen*) und physischen Ereignissen (Hirnzuständen) gibt. Warum ist es trotzdem nicht möglich, strikte Gesetze (nach dem Vorbild physikalischer Gesetze) herauszufinden, die die beiden Bereiche miteinander verknüpfen? Sind mentale Phänomene letztlich anomische Entitäten, die überhaupt nicht durch gesetzesartige Verallgemeinerungen erfaßt werden können?

Wenn die Frage nach der **Intentionalität des Mentalen** als die Frage nach seinem intentionalen Gehalt gelesen werden kann, dann kann man also das **Leib-Seele-Problem** als die Frage nach dem *kausalen* Gehalt mentaler Zustände verstehen. Die meisten Menschen glauben, daß ihre Meinungen, Wünsche, Gefühle oder Gedanken *als so beschriebene* eine direkte Wirkung auf ihren Körper, ihr Verhalten und ihre Handlungen ausüben. Mentale Zustände sind also Phänomene, die nach landläufiger Meinung einen Wirkungsgehalt haben: Sie können als *Ursachen* physischer Prozesse in Erscheinung treten. Zu klären, ob und wie dies möglich sein kann, ist die Aufgabe der modernen philosophisch-metatheoretischen Diskussion des

Leib-Seele-Problems. Sie ist durch starke Interdisziplinarität und die Befruchtung durch neue empirische Erkenntnisse charakterisiert. Dasselbe gilt auch für die Bemühungen um eine philosophische Theorie des Subjekts geistiger Zustände.

Subjektivität. Man kann sagen, daß die **dritte** Klasse von philosophischen Problemen für eine moderne Theorie des Geistes nicht durch die epistemische und kausale Einbettung des Mentalen in die Natur entsteht, sondern durch die Einbindung des Geistigen *in sich selbst*. Insofern sind diese Probleme aus naturalistischer Perspektive auch die schwierigsten: Für sie ist die Innerlichkeit subjektiven Bewußtseins maßgebend. Wie aber soll eine am Objektivitätsideal der empirischen Wissenschaften ausgerichtete Theorie des Geistes – unter Zugriff auf öffentliche und intersubjektiv kontrollierbare Quellen der Erkenntnis – sich diesem Merkmal des Geistigen annähern? Gibt es überhaupt eine Hoffnung, ein Phänomen wie *Bewußtsein* einer erfolgversprechenden naturalistischen Analyse zuzuführen – ohne sich automatisch der Produktion von Kategorienfehlern schuldig zu machen oder in empiristische Zirkelschlüsse zu geraten?

Man kann die philosophische Frage nach der **Subjektivität des Mentalen** auch als ein Problem zweiter theoretischer Ordnung bezüglich psychischer Prozesse formulieren und dabei wiederum von einer Form von *Gehalt* sprechen: Wie kommt es, daß mentale Zustände einen *phänomenalen* Gehalt besitzen? Denn mentale Zustände sind *qualitative* Zustände – man denke nur an die Schmerzhaftigkeit von Schmerzen oder die subjektive Qualität der Röte, die eine Rotwahrnehmung begleitet. Das Problem besteht darin, daß eine große Anzahl geistiger Vorgänge einen spezifischen Erlebnisgehalt besitzen, der nur derjenigen Person zugänglich ist, deren innere Prozesse sie sind. Es scheint, als wäre ein wissenschaftlicher Zugriff auf solche privaten Aspekte des Mentalen prinzipiell unmöglich. Wenn das so ist, dann kann es aber auch niemals eine vollständige und befriedigende wissenschaftliche Beschreibung dieses Bereichs der Wirklichkeit geben. In diesem Gedanken wird die generelle, über die Philosophie des Geistes hinaus verweisende Virulenz des Problems deutlich. Auch aus diesem Grund stellt das Problem der **Subjektivität mentaler Zustände** das thematische Zentrum der vorliegenden Arbeit dar. Ich werde es deshalb in diesem ersten Kapitel einer näheren Betrachtung unterziehen. Im weiteren Verlauf werde ich dann versuchen, vor dem Hintergrund einer naturalistischen Theorie mentaler Repräsentation einige Vorschläge zu seiner Lösung zu entwickeln.

Es scheint, als habe sich die fruchtbare metaphysische Phase der Entwicklung des philosophischen Geistbegriffs ihrem Ende zugeneigt. Nach einer Phase intensiver sprachanalytischer Beiträge, die nicht selten von platten ontologischen Reduktionismen begleitet wurden, kündigt sich nun die Entwicklung einer naturalistischen Metatheorie mentaler Phänomene an. Sowohl die Einzeldisziplinen als auch die Philosophie sind heute bereit, das Mentale ernstzunehmen als eines der faszinierendsten und rätselhaftesten Naturphänomene, das wir kennen. Das Mentale ist nun kein Gegen-

stand von idealistischer Spekulation oder impressionistischer Begriffsdichtung mehr und es wird auch nur noch selten zum Scheinproblem degradiert. Die empirischen Neuro- und Kognitionswissenschaften haben ihre Aufmerksamkeit mit nie gekannter Intensität dem menschlichen Gehirn, seinen kognitiven Kapazitäten und den von ihm instantiierten psychologischen Eigenschaften zugewendet. Gleichzeitig hat es von Seiten der Philosophie des Geistes Bereicherungen unseres Begriffsapparates gegeben und auch andererseits Vorschläge, das mentalistische Alltagsidiom vollständig zu eliminieren. Das zeigt, daß die begriffsanalytische Begleitung der Einzeldisziplinen sowohl in der Bereicherung des oben bereits angesprochenen Begriffsnetzes durch zusätzliche oder besser plazierte, als auch in der Lösung einzelner – im Extremfall *aller* – kategorialer Knoten bestehen könnte. Karl Popper hat einmal gesagt, daß wir mit der Wirklichkeit immer genau in den Momenten in Kontakt sind, in denen unsere Theorien an ihr scheitern. In diesem Sinne besteht Philosophie des Geistes auch in dem Versuch, den Realitätskontakt unserer Theorie des Mentalen zu optimieren, indem man unnötige Spannungen innerhalb des über die – sich stetig entwickelnde – Datenlandschaft der empirischen Psychologie geworfenen Begriffsnetzes zu verhindern sucht. Das kann manchmal durch das Lösen obsolet gewordener theoretischer Knoten erreicht werden. Die Philosophie muß dabei jedoch gleichzeitig subtilere Verknüpfungen in Bereichen verlorenen Wirklichkeitskontakts anbieten und auf diese Weise versuchen, eine logische Topographie der aus neuen empirischen Erkenntnissen erwachsenden Problemlandschaft zu erarbeiten. Auf diese Weise kann sich das konzeptuelle Netz enger an die nun deutlicher hervortretende Landschaft aus empirischen Daten anschmiegen. Und so ergeben sich neue Möglichkeiten, unsere Theorie des Mentalen an der inneren Wirklichkeit scheitern zu lassen.

1.2 Subjektivität als Kernstück einer modernen Theorie des Geistes: Die drei Hauptelemente des Problems

Wir sind Wesen, die einen Teil ihrer eigenen Zustände bewußt erleben. In dieser knappen Formulierung sind bereits die drei wesentlichen Elemente des psychologischen Subjektbegriffs benannt: Identität, Bewußtsein und Erlebnisträgerschaft. Diese drei begrifflichen Elemente beziehen sich auf die drei phänomenologischen Säulen psychologischer Subjektivität, also auf das Fundament und die tragenden Elemente des phänomenalen Ich. Die erste phänomenologische Säule ist das **Selbst**. Es vereint alle subjektiven Zustände als *meine* Zustände und wird gleichzeitig durch sie konstituiert. Die zweite phänomenologische Säule des Ich ist das **Bewußtsein**: Die scheinbar unmittelbare und direkte *Gegebenheit* seiner Inhalte. Die Inhalte des Selbstbewußtseins jedoch besitzen zudem noch ein weiteres Charakteristikum, eine **Erlebnisqualität**. Diese Qualität des Erlebens bildet die dritte Säule des phänomenalen Ich, sie ist dasjenige Merkmal, welches unserem

Innenleben seine unhintergehbare Konkretheit, seinen Reichtum und seine Buntheit verleiht. Die eben genannten begrifflichen Elemente von Subjektivität – Identität, Bewußtsein und Erlebnisträgerschaft – versuchen, das Wesen der drei phänomenologischen Säulen unseres selbstbewußten Erlebens zu erfassen und einzufangen. Sie tun dies in einer sehr vagen und häufig äquivoken Art und Weise, die es schwierig macht, den Begriff des psychologischen Subjekts als den Begriff eines *Explanandums* zu verstehen: Wenn wir nach wissenschaftlichen Erklärungen oder der korrekten philosophischen Analyse des Phänomens „Subjektivität" – im Sinne des in der Einleitung angesprochenen, diese Erklärungen miteinander verbindenden metatheoretischen Kommentars – suchen, dann wissen wir im Grunde überhaupt nicht, wonach wir suchen sollen.

Dennoch gibt es die oben angesprochenen drei Hauptaspekte des Problems. Darum werde ich nun der Reihe nach einen Blick auf jeden dieser drei phänomenologischen Hauptaspekte werfen. Mein Ziel wird dabei sein, zu einem etwas klareren vorläufigen Bild der Problemlandschaft zu gelangen. Im vierten Kapitel dieses Buches werden wir dann aus einer anderen Richtung etwas weiter in diese Landschaft eindringen, indem wir prominente Beiträge der Gegenwartsphilosophie zu ihrer Erkundung untersuchen. Vorerst jedoch werde ich einen Katalog der Rätsel anbieten, die wir mindestens lösen müssen, wenn wir an einem theoretischen Verständnis unseres eigenen Subjektseins interessiert sind. In den beiden auf diese einführenden Bemerkungen folgenden Kapiteln werde ich dann einzelne Partien des philosophischen Puzzles näher untersuchen und dabei die theoretischen Mittel entwickeln, mit deren Hilfe ich am Ende dieser Arbeit schließlich einige Lösungsmöglichkeiten aufzeigen werde. Im Gang der Überlegungen wird deutlich werden, daß der psychologisch-philosophische Subjektbegriff durch das Entstehen der neuen Theorie des Geistes vor einer Anzahl tiefgreifender semantischer Transformationen steht. Mit ihnen wird sich auch die ihn traditionell umgebende Landschaft schwieriger philosophischer Fragen deutlich verwandeln – wobei sie weitgehend zu einer *terra incognita* werden wird, die sich uns nur erschließt, wenn wir uns selbst besser verstehen. Wir befinden uns also in einer neuen und schwierigen Situation. Was genau ist es, das das Unternehmen einer postmetaphysischen Theorie der Subjektivität zu einer so schwierigen und gefährlichen Expedition in ein unbekanntes Gebiet macht?

Zuallererst ist die Frage nach dem Subjekt psychischer Zustände immer auch die Frage nach dem Selbst. Das ist darum so, weil die dieses Subjekt konstituierenden inneren Erlebnisse auf rätselhafte Weise immer seine sind. Ich bin ein Wesen, das eine psychische Identität besitzt – ein Wesen, das seine bewußten, inneren Erlebnisse in einem sehr direkten und unvermittelten Sinn als seine eigenen erfährt. Ein Schmerz ist nicht einfach schmerzhaft, er ist *mein Schmerz*. Auch das Überzeugtsein von einer philosophischen oder wissenschaftlichen Theorie schwebt nicht einfach frei in einem inneren Bewußtseinsraum, es ist *meine Überzeugung*. Mentale Zustände besitzen also ein sehr prägnantes relationales Charakteristikum.

Dieses relationale Merkmal phänomenaler Zustände kann man aus der Perspektive der ersten Person (die sie durchlebt und deren Zustände es sind) als ihre *„Meinigkeit"* bezeichnen. Mein bewußtes Leben baut sich aus einer ständigen und unvorhersehbaren Verkettung von Zuständen der Welt auf und ein Teil dieser Zustände sind *meine* Erlebnisse. Diese Serie meiner Erlebnisse wird scheinbar ganz ohne mein eigenes Zutun zu einer solchen, denn meine subjektiven Erlebnisse bilden wie die Perlen einer Kette eine neue, größere Gestalt. Diese Gestalt bin ich *selbst,* und die Gestalt aus Erlebnisperlen verleiht mir auch durch ihre Geschichtlichkeit ein Gefühl von Identität (eine zeitliche Gestalt). Ob es einen Faden gibt, der all die Perlen zusammenhält, weiß ich nicht – aber vor aller philosophischen Reflexion scheint eines gewiß: Jede einzelne Perle auf der Erlebniskette *gehört zu mir.* Ich selbst kann und muß dazu nichts tun. Es scheint nichts Gewisseres in der Welt zu geben, als daß meine Erlebnisse *meine* Erlebnisse sind.

Aus der externen Perspektive werde ich dieses eben angesprochene relationale Merkmal unserer inneren Zustände ab jetzt auch als ihre „Subjektgebundenheit" oder ihre „Subjektzentriertheit" bezeichnen: In Standardsituationen sind mentale Zustände immer die Zustände eines phänomenalen Ich. Es ist dieser strukturelle Aspekt unseres Innenlebens, auf den ich mich in dieser Arbeit konzentrieren werde, weil er, wie ich glaube, den Kern des philosophischen Problems darstellt. Das eigentliche phänomenologische Rätsel ist nämlich die Tatsache, daß wir ein *zentriertes Bewußtsein* besitzen – einen um einen Brennpunkt herum aufgebauten inneren Erlebnisraum.

Diesem phänomenologischen Kern des Rätsels entspricht eine logische Problemdimension: Es hat den Anschein, als ob sich die Subjekt-Argumentstelle[1] aus Beschreibungen mentaler Zustände prinzipiell nicht eliminieren läßt. Diese semantische Annahme liegt auch einem einflußrei-

[1] Ulrich Blau hat die Paradoxien untersucht, die durch das Selbst als fiktivem Fixpunkt in unserer fließenden Ontologie erzeugt werden. Er lokalisiert das Grundproblem in dem für uns unhintergehbaren Mechanismus der Objekt- und Begriffsbildung, der ständig neue Objekte erzeugt, die sich bei genauem Hinsehen als unterbestimmt erweisen, d. h. ihre scharfe Identität einbüßen. Besonders interessant erscheint mir sein Hinweis, daß das Proto-Objekt der klassischen Logik eben genau das *Selbst* ist: *„Solange die Mathematik die klassische Identitätslogik verwendet, kommt sie von ihrem ersten Bild nicht los: dem scharf individuierten Objekt. Das Urbild bin ich selbst, und praktisch komme ich nicht davon los,* E [der **Essentialist**; Anmerkung TM] *hat recht. Bei kritischer Betrachtung werde ich unscharf und löse mich auf, S* [der **Skeptiker**; Anmerkung TM] *hat recht. Aber in was löse ich mich auf? Seine Objekte sind bei kritischer Betrachtung genauso fiktiv, und praktisch ist er genauso real wie E. Die Paradoxie hat keine natürliche Lösung."* (Blau 1986: 194) Blau zeigt ebenfalls, daß selbst beim Übergang zur Betrachtung absolut scharf individuierter Objekte – also zu Mengentheorie – der andauernde Transformationsprozeß der Objekt- und Begriffsbildung ein Grundmerkmal der *Perspektivität* nicht verliert: *„Viele Mengenbezeichnungen und alle echten Klassenbezeichnungen sind nur eindeutig relativ zum kontextuellen Mengenuniversum, das immer wieder erweitert werden kann und muß."* (Blau 1986: 193, dazu auch 1985: 369–459, 1986b) Dieser Gedanke erscheint mir von großer Bedeutung in Zusammenhang mit der Nagelschen These bezüglich der Existenz von *perspectival facts*, vgl. Kapitel 4.

chen Argument gegen die Identitätstheorie (als einer Theorie bezüglich der Relation zwischen physischen und psychischen Phänomenen) zugrunde, das Saul Kripke entwickelt hat.[2] Dieses Argument – welches ein breites Echo in über fünfzig Aufsätzen hervorgerufen hat – greift in seinem ersten Schritt die von Vertretern früher, postbehavioristischer Leib-Seele-Theorien[3] behauptete Kontingenz der Identitätsrelation zwischen psychischen und Hirnprozessen an, indem darauf hingewiesen wird, daß die in ihr eingesetzten Terme starre Designatoren *(„rigid designators")* sind. Das heißt: Starre Bezeichnungsausdrücke wie *„mein Gehirnvorgang vom Typ A"* und *„mein Schmerz B"* besitzen in allen möglichen Welten dieselben Referenten. Sie funktionieren wie Eigennamen, weil sie ihre Referenzobjekte durch eine *essentielle Spezifikation* erfassen. Keine Angst – ich möchte das Kripkesche Argument hier nicht diskutieren, sondern lediglich die Aufmerksamkeit meiner Leser auf die Tatsache lenken, daß es genau der durch die Analyse von mentalistischen Beschreibungen des Typs *„mein Schmerz vom Typ A"* als rigide Designatoren ausgedrückte semantische Essentialismus ist, der ein wichtiges Merkmal unseres Erlebens in Form der aus ihm entstehenden Intuition wiedergibt, daß meine mentalen Zustände immer schon *meine* mentale Zustände sind und sein müssen. Wer glaubt, es könnte innere Erlebnisse geben, die *niemandes* Erlebnisse oder nur zufällig die inneren Zustände dieses bestimmten psychologischen Subjekts sind, der hat – so Kripke – überhaupt nicht verstanden, über was für Zustände wir in solchen Fällen sprechen. Wenn das aber so ist, dann können grundsätzliche Zweifel an der Theoriefähigkeit von Subjektivität aufkommen: Wenn die Subjekt-Argumentstelle sich aus Beschreibungen mentaler Zustände prinzipiell nicht eliminieren[4] läßt, dann können solche Beschreibungen auch nicht nomologisch mit objektiven Aussagen (etwa über Vorgänge im Gehirn) verknüpft werden, weil sie zu einem ganz anderen *Typ* von Beschreibung gehören.

Die Frage nach der Subjektivität mentaler Zustände gehört in das theoretische Feld des Leib-Seele-Problems. Je nachdem, wie man die vielen Einzelfragen dieses schwierigen und komplexen Bereichs der Theorienbildung gewichtet, kann man sie auch als das zentrale Problem schlechthin konstruieren: Was die meisten von uns – auch diejenigen mit starken materialistischen Neigungen – nicht verstehen können, ist, wie unser *eigenes* Bewußt-

[2] Vgl. Kripke 1971, Kripke 1972. Ich habe dieses Argument und die Art und Weise, in der cartesianische Intuitionen in es einfließen, an anderer Stelle diskutiert. Vgl. Metzinger 1985: 213ff und auch die Kritik von Bartels an Kripke in Bartels 1984. Eine gute Rekonstruktion gibt Lycan 1987, Kapitel 2; vgl. auch Stegmüller 1986: 312ff.

[3] Vgl. etwa Armstrong 1965, 1968, Feigl 1960, 1967, Lewis 1966(deutsch 1989), Lewis 1972 (deutsch 1989), Place 1956, 1988, Smart 1959, 1961, 1962, 1963a, 1963b; Erläuterungen finden sich in Borst 1970: 13ff, Bieri 1981: 36ff und Metzinger 1985: 16ff und 210ff.

[4] Eine sehr luzide Diskussion der quasi-logischen „Gesetze der Subjektivität" und der mit ihnen verbundenen Nicht-Eliminierbarkeit des subjektiven Standpunkts bietet Colin McGinn an; vgl. McGinn 1983, besonders Kapitel 3 und 6.

sein mit der scheinbar direkten Gegebenheit, Homogenität und Konkretheit seiner Inhalte nichts als das Resultat von Informationsverarbeitungsprozessen in unserem zentralen Nervensystem sein kann. Wenn der intentionale Gehalt mentaler Repräsentate ihr wesentliches Merkmal ist, dann scheint es *prima facie* unverständlich, wie dieser Gehalt durch eine natürliche Genese in die Welt gekommen ist. Es wird auch unklar, wie er – zum Beispiel durch Korrelationen mit gesetzesartigem Charakter[5] – mit anderen Ereignissen in der Welt verknüpft sein kann, mit Ereignissen, die wir unter neurobiologischen, funktionalen oder physikalischen Beschreibungen kennen. Wenn zudem *phänomenaler* Gehalt eine höhere und theoretisch relevantere Eigenschaft psychischer Zustände sein sollte als dieser repräsentationale Gehalt, dann wird man **erstens** mit der Frage nach der Beziehung zwischen zwei Arten von mentalem Gehalt konfrontiert und **zweitens** mit einer neuen Variante des Leib-Seele-Problems. Sie lautet: Wie können in einer objektiv völlig mittelpunktlosen Welt *phänomenale Standpunkte* entstehen? Wie können Biosysteme in einer objektiv perspektivelosen Umwelt eine interne Repräsentation dieser Welt erzeugen, die essentiell *perspektivisch* ist? Wie können Gehirne qua modular organisierte und Information weitgehend parallel-distribuiert speichernde und verarbeitende Systeme ein *zentriertes Bewußtsein*[6] erzeugen – ein Bewußtsein, das um einen phänomenalen Brennpunkt herum aufgebaut ist?

Wir wissen, daß der Einsatz eines ontologischen Mittels zur Lösung des Problems ein theoretischer Holzweg ist, weil er uns – von Descartes bis Popper – zu einer Beschreibung des Modus psychophysischer Interaktion führt, der gezwungenermaßen auf Metaphern und Analogien rekurrieren muß. Er zwingt uns zu der Annahme einer transphysikalischen Kausalität, die von lokalen Verletzungen des Energieerhaltungssatzes im Gehirn aus-

[5] Der Name Donald Davidsons ist mit der These von der Anomalie des Mentalen verknüpft, die besagt, daß mentale Ereignisse *als so beschriebene* niemals Gegenstand strikter, homonomer Verallgemeinerungen nach dem Vorbild physikalischer Gesetze sein können, da die mentalistische Beschreibungsebene andere logische Subjekte einsetzt als die physikalische oder neurophysiologische (mentale Zustände sind Personenzustände). Es kann also keine psychophysischen Gesetze geben – die Kluft zwischen mentalen und physischen Ereignissen ist prinzipiell unüberbrückbar, weil wir aus prinzipiellen Gründen die Glieder von psychophysischen Kausalketten nicht *en détail* spezifizieren können. Die Pointe der Davidsonschen These (die zur Entwicklung der verschiedenen Formen der Supervenienz-Theorie geführt hat) liegt aber nicht in der „Gesetzlosigkeit" und der *nomologischen Inkommensurabilität* des Mentalen. Sie liegt in der Konklusion, daß jedes mentale Ereignis – wenn es ein physikalisches Ereignis verursacht – unter eine physikalische Beschreibung fällt, weil die Verursachung unter irgendeiner Beschreibung ein striktes Gesetz instantiieren muß und nur physikalische Gesetze Kandidaten für strikte Gesetze sind. Die so beschriebene Position des anomalen Monismus behauptet also gleichzeitig die Irreduzibilität mentaler Ereignisse *qua* mentaler Ereignisse und ihre *Identität* mit physikalischen Ereignissen. *„Auch wenn es jemanden gäbe, der die gesamte physikalische Geschichte der Welt kennt, und jedes mentale Ereignis mit einem physikalischen Ereignis identisch wäre, würde nicht geschlossen werden können, daß er auch nur ein einziges mentales Ereignis (natürlich als so beschriebenes) voraussagen und erklären kann."* (Davidson 1981: 89) Vgl. Davidson 1970 (1981), Metzinger 1985: 97ff, Stanton 1975.

[6] Vgl. Metzinger 1989.

geht, und verwickelt uns in Schwierigkeiten bei der Formulierung von psychophysischen Gesetzen.

Psychologische Subjektivität entsteht dadurch, daß zum intentionalen Gehalt mentaler Repräsentate eine Relation hinzutritt. Die physische Realisierung dieser Relation erst macht interne Repräsentationsprozesse in informationsverarbeitenden Systemen zu *subjektgebundenen*: Aus einem Beinschmerz wird ein Schmerz in *meinem* Bein; aus einer propositional repräsentierten positiven Wahrscheinlichkeitsaussage wird *mein Glauben, daß p*; aus einer repräsentational isolierten Verlustangst wird *meine* Eifersucht und so weiter. Was uns also an dieser Stelle interessieren muß, ist das zweite Relatum der Subjektbeziehung. Um Subjektivität als ein psychisches Phänomen verstehen zu können, brauchen wir folglich eine naturalistische Theorie des Ich. Eine naturalistische Theorie wird sich dadurch auszeichnen, daß mit ihr ein Erkenntnisfortschritt möglich ist: Die Innerlichkeit unserer Innenperspektive werden wir nur dann besser verstehen, wenn wir es mit Hilfe empirisch gehaltvoller und deshalb auch falsifizierbarer Theorien tun. Es mag durchaus sein, daß Subjektivität und Selbstbewußtsein die großen Rätsel bleiben, an denen das wissenschaftliche Erkenntnisideal zunichte wird, transzendentale Mysterien, die den mit ihm verknüpften monistischen Optimismus endgültig scheitern lassen. Andererseits ist es offensichtlich, daß spekulativ-idealistische Ansätze obsolet geworden sind, weil sie eben jenen Erkenntnisfortschritt nicht produzieren können. Was wir benötigen, ist eine naturalistische Theorie des Selbst, die uns überzeugend erklärt, wie von unserem Gehirn erzeugte innere Repräsentate von Teilen der Welt zu *unseren* mentalen Erlebnissen werden können, ohne dabei mit empirischen Fakten aus den Kognitions- und Neurowissenschaften zu kollidieren.

An den ethischen Aspekten einer Subjektivitätstheorie kann man die Bedeutung der Subjektzentriertheit des Mentalen weiter verdeutlichen. Die Tatsache, daß in einem gegebenen biologischen System (z. B. einem im Netz zappelnden Fisch) interne Zustände auftreten können, die denselben *kausalen* Gehalt (also einen analogen Einfluß auf das behaviorale Profil des Fischs) und denselben *repräsentationalen* Gehalt haben wie entsprechende Typen von mentalen Zuständen – zum Beispiel: „Angst", „Panik", „Verzweiflung" – beim Menschen, qualifiziert es nämlich noch nicht automatisch als moralisches Objekt, als Gegenstand moralischen Handelns. Erst wenn diese Zustände auch *seine* sind, wenn sie im phänomenalen Modus der „Meinigkeit" die innere Bühne betreten, werden sie relevant für Fragestellungen wie die der moralischen *Inter*subjektivität.

Phänomenale „Meinigkeit" scheint eine Schlüsselrolle für das Problem der Subjektivität des Mentalen einzunehmen, da diese Eigenschaft aufs engste verknüpft ist mit dem Erlebnischarakter innerer Zustände. Vielleicht kann man diese Eigenschaft durch eine Relation zwischen unterschiedlichen Typen von mentalen Repräsentaten erklären. Gelänge dies auch für das Phänomen der Subjektivität als Ganzem, dann könnte man auf begrifflicher Ebene von dem gelungenen Projekt einer Reduktion von

Subjektivität auf Repräsentation sprechen. Wie ich bereits angedeutet habe, könnten erst vor dem Hintergrund einer solchen Theorie für die Ethik so immens bedeutsame Fragen entschieden werden wie die, welche Arten von informationsverarbeitenden Systemen etwa leidensfähig sind, Todesangst erleben können und so weiter. Dann jedoch könnten, um mit Frank zu sprechen, Leibnizens „träumende" und „fühlende Monaden" eine nicht nur romantische Realität gewinnen.[7] Diese Frage scheint von großer Bedeutung für eine nicht-speziezistische Ethik. Sie kann aber nur vor dem Hintergrund einer überzeugenden Theorie der Subjektivität entschieden und für die ethische Diskussion fruchtbar gemacht werden.

Die Fragen nach psychischer Identität, qualitativem Selbstbewußtsein und der Subjektgebundenheit mentaler Zustände bilden, wie gesagt, den Kern des Problems und den Schwerpunkt der vorliegenden Arbeit. Darum ist es nicht nötig, an dieser Stelle mehr zu diesem **ersten** Aspekt des Problems zu sagen – weil seine Struktur im weiteren Verlauf unserer begrifflichen Exkursion immer deutlicher zutage treten wird. Es ist aber unmöglich, sich dieser Struktur ohne ein Verständnis der mysteriösen Tatsache zu nähern, daß subjektzentrierte Erlebnisse durch Bewußtseinsinhalte gebildet werden.

Denn es gibt **zweitens** das Problem des *Bewußtseins*: Ich erlebe meine inneren Zustände und auch die der Welt als Bewußtseinsinhalte. Das bedeutet: Diese Inhalte werden durch eine gemeinsame Qualität miteinander verbunden, eben durch ihre subjektiv erlebte Bewußtheit. Analysiert man „bewußt" als ein zweistelliges Prädikat, dann wird man auf das eingangs erwähnte Problem der Intentionalität des Mentalen gestoßen. In dieser Form ist Bewußtsein immer Bewußtsein *von etwas*. Interpretiert man „bewußt" als einstelliges Prädikat, so verweist es allem Anschein nach auf eine primitive und nicht weiter analysierbare Eigenschaft mentaler Zustände. Diese Qualität scheint sich, wie wir gleich sehen werden, dem Instrumentarium wissenschaftlicher Beschreibung aus prinzipiellen Gründen zu entziehen. Andererseits ist sie von großer theoretischer Bedeutung, zum Beispiel weil sie traditionell (seit der „cartesianischen Katastrophe"[8]) das Kriterium für eine exklusive Unterscheidung zwischen physischen und mentalen Phänomenen liefert. Sie bildet nämlich den Kernaspekt eines weiteren wichtigen philosophischen Problems: des Leib-Seele-Problems. Wir wissen heute einerseits, daß das Auftreten von Bewußtheit eng mit neurobiologischen

[7] Vgl. Frank 1991: 250.

[8] Kathy Wilkes hat das Auftreten des Terms „Bewußtsein" in der abendländischen Geistesgeschichte, seinen Ursprung in erkenntnistheoretischen Interessen und der endgültigen Überwindung der aristotelischen Denkfigur der *psyche* durch die cartesianische Revolution untersucht. Sie argumentiert dafür, daß das Schisma zwischen „bewußt" und „unbewußt" eine große Menge der wissenschaftlich interessanten Phänomene unklassifiziert läßt und dafür, daß „Bewußtsein" niemals aus der *common sense*–Taxonomie in die wissenschaftliche übergehen wird, weil der Begriff sich auf keine natürliche Art bezieht. Vgl. Wilkes 1988, und auch 1984.

Ereignissen korreliert, andererseits scheint eine physikalistische Reduktion oder eine funktionale Analyse dieses Phänomens unmöglich.

Selbst wenn es uns gelänge, einer Maschine all die funktionalen Eigenschaften zu verleihen, die wir bei einem *Menschen* als Merkmale von Bewußtheit ansehen, würden wir deshalb noch lange nicht glauben, daß diese Maschine auch bewußt *ist*. Angenommen, wir programmieren ein künstliches System auf eine Weise (oder bringen ihm bei, wie es sich selbst programmiert), die es ihm ermöglicht, sich intelligent und zielgerichtet zu verhalten und sich erfolgreich in der Welt zu bewegen – wer von uns würde glauben, daß wir *Künstliches Bewußtsein* realisiert hätten und nicht bloß *Künstliche Intelligenz?* Wären wir mit einem natürlichsprachigen System konfrontiert, welches auf eine derart umfangreiche Wissensbasis zugreifen könnte und über ein so reiches implizites Hintergrundwissen bezüglich unserer Lebensform verfügen würde, daß es den Turing-Test[9] bestünde und für beliebige menschliche Kommunikationspartner nicht mehr aufgrund von Defiziten in der Fähigkeit, ein intelligentes Gespräch mit ihnen zu führen, als künstliches System zu entlarven wäre: Hätten wir nicht immer noch starke Zweifel daran, daß in einem solchen künstlichen System das Licht der Bewußtheit scheint? Wenn ein sprachbegabter Roboter der Zukunft durch ein so komplexes und subtiles Netz von Kausalketten und funktionalen Zuständen mit seiner Umwelt verwoben wäre, daß wir bei der Vorhersage und Erklärung seiner Handlungen nicht umhin könnten, ihn als *intentionales System*[10] zu beschreiben, als ein solches System, dessen innere Zustände wir als Zustände mit intentionalem Gehalt analysieren müssen, wenn wir seine Verhaltensmuster überhaupt noch verstehen wollen – würden wir deshalb annehmen, daß wir nun auch ein *phänomenales System* konstruiert haben?

Sicher nicht. Wir können vielleicht kognitive Agenten konstruieren, handelnde Systeme, die aktiv Wissen über die Welt aufbauen und für gewisse Zwecke einsetzen. Es mag auch sein, daß solche technischen Dämonen uns recht bald durch die Komplexität ihres behavioralen Profils und durch den repräsentationalen Reichtum ihrer internen Zustände verblüffen werden. Aber solange wir diese internen Zustände durch Turing-Maschinentafeln, Blockschaltbilder oder Flußdiagramme erläutern können, solange können wir immer sagen: „All dies ist vollkommen *ohne* Bewußtsein möglich! Wir haben es mit einem mechanischen System zu tun, alle seine Eigenschaften sind aus den Eigenschaften seiner Teile und ihren Beziehungen untereinander abzuleiten. Selbst wenn es dem System gelingt, durch sein enormes Wissen, seine Flexibilität und seine beeindruckenden Fähigkeiten auf der

[9] Vgl. Turing 1950; dazu auch Beckermann 1988, 1990, Boden 1990a, Davidson 1990, Narayan 1990, Putnam 1987, Rheinwald 1991, Searle 1980, 1986. Alan Turings bahnbrechender Aufsatz über berechenbare Zahlen und das Entscheidungsproblem (Turing 1937) darf als der theoretische Eintritt der Menschheit in das Computerzeitalter gelten. Einen guten Überblick über die Entwicklung in dem seitdem vergangenen halben Jahrhundert gibt der von Rolf Herken zusammengestellte Sammelband; vgl. Herken 1988.

[10] Vgl. Dennett 1981, 1987.

Performanzebene unsere Intuitionen ins Wackeln zu bringen, dürfen wir eines nicht vergessen: Das System wird niemals wirklich verstehen, was der Unterschied zwischen Wachen und Schlafen ist . . .“

Unangenehm würde die Situation erst, wenn einer unserer künstlichen kognitiven Agenten (in der ihm eigenen *sachlichen* Art und Weise) erwidert: „Aber ihr wißt es ja *selbst* gar nicht!“ Wenn ein künstliches System uns in eine Diskurssituation zwingt und rational für seine *eigene* Theorie des Geistes zu argumentieren beginnt, dann wären wir gezwungen, ihm zu zeigen, daß der Term „Bewußtsein“ einen sinnvollen Platz in einer wissenschaftlichen Taxonomie psychischer Zustände einnimmt, die der seinigen überlegen ist. Sollte uns dies nicht gelingen[11], müßten wir uns auf außerwissenschaftliche Erkenntnisquellen berufen und sagen: „Was Bewußtsein ist, ist *evident*. Alle bewußten Wesen – zu denen Du nicht gehörst – *wissen*, daß sie bewußt sind und verstehen sofort, was mit diesem Ausdruck gemeint ist.“ Bestimmt würden an diesem Punkt eine Reihe materialistischer Philosophen die Solidarität aller biologischen Wesen durchbrechen und mit dem Stolz der Ketzer verkünden, daß auch *sie* noch nie verstehen konnten, was es eigentlich heißt, daß *sie* „Bewußtsein“ besitzen sollen. Wollte man eine solche unheilige anti-cartesianische Allianz erfolgreich zurückschlagen, so müßte man überzeugend darlegen können, daß das, was wir alle immer schon ganz selbstverständlich als unser Bewußtsein bezeichnen, ein reales Phänomen ist, dem eine präzise Position innerhalb einer naturalistischen Theorie des Geistes zugewiesen werden kann.

Dies mag sich als ein schwieriges Projekt herausstellen. Wenn wirklich ist, was *wirkt*, dann könnte sich herausstellen, daß Bewußtsein bestenfalls ein Epiphänomen[12] ohne kausale Wirksamkeit ist. Denn kausale Erklärun-

---

[11] Wenn es uns (bereits vorher) gelingt, dann werden wir diese wissenschaftliche Einsicht in das Wesen der Bewußtheit prinzipiell auch technologisch umsetzen können. Karl Popper verleiht dem mit dieser Einsicht verbundenen konservativen Unbehagen Ausdruck, indem er schreibt: *„Turing (1950) sagte es einmal ungefähr so: Gib genau an, worin deiner Meinung nach ein Mensch einem Computer überlegen sein soll, und ich werde dir einen Computer bauen, der deinen Glauben widerlegt. Wir sollten Turings Herausforderung nicht annehmen, denn jede hinreichend genaue Bestimmung könnte prinzipiell zur Programmierung eines Computers verwendet werden.“* (Popper/Eccles 1982: 257)

[12] Eine gute Analyse der Problematik bietet Peter Bieri in *„Trying out Epiphenomenalism“* an; vgl. Bieri 1990. Eines der zentralen Probleme ist in diesem Zusammenhang die Frage, ob der intentionale Gehalt mentaler Zustände (bzw. ihre fragliche Bewußtheit) eine kausale Rolle bei der Genese von Verhalten spielen muß, um dieses in manchen Fällen als *rationales Verhalten* bezeichnen zu können. Mir scheint, daß wir, um noch wirklich an unsere eigene Vernünftigkeit als auch introspektiv erlebtes Phänomen glauben zu können, dezidierte Realisten bezüglich der Rationalität mancher unserer inneren Operationen sein müßten. Bieri glaubt dies nicht, weil für ihn Rationalitätszuschreibungen auf abstrakte Entitäten Bezug nehmen (vgl. hierzu Dennett 1987) und mit einem empirisch-methodologisch fundierten Epiphänomenalismus verträglich sind. *„On the contrary, rationalizing explanations proceed and can succeed independently of any questions regarding internal mechanisms. It is not the causal properties of reasons which are invoked but the content properties pertinent to the possibility of justifying behavior. Thus, it doesn't matter that rationalizing explanations are not extendible like causal explanations. They are not supposed to be (. . .). So epiphenomenalism does not endanger psychological understanding as a rationalistic hermeneutics.“* (Bieri 1990: 34)

gen menschlicher Handlungen vom neurowissenschaftlichen Typ müssen nicht auf Beschreibungen innerer Zustände rekurrieren, in denen das Prädikat „bewußt“ auftaucht. Bei unserer Erforschung des menschlichen Gehirns entdecken wir immer neue Beschreibungsebenen: Physikalische, neurochemische, neurobiologische, neuropsychologische, funktionale, informationale und repräsentationale. Jede dieser Ebenen besitzt ihre eigenen logischen Subjekte, denen wir bei unseren Erklärungen unterschiedliche Eigenschaften zuweisen. Mit der Expansion unseres empirischen Wissens nimmt die Zahl der Beschreibungsebenen zu, die ihnen zugeordneten Sets von Prädikaten werden stetig subtiler und differenzierter – jedoch taucht das Prädikat „bewußt“ niemals auf. Aber immer, wenn wir ein biologisches, funktionales oder repräsentationales Subsystem der gigantischen neuronalen Maschinerie etwas besser verstanden haben, taucht wieder dieselbe Frage auf: *Aber wäre all dies nicht auch ohne Bewußtsein möglich*?

Es mag durchaus sein, daß die zukünftige Theorie des Geistes in weiten Teilen stark kontraintuitiv und damit emotional unbefriedigend sein wird. Vielleicht wird es niemals gelingen, das vom Gehirn selbst gezeichnete Bild seiner eigenen Aktivität und das von der Wissenschaft entworfene Bild derselben miteinander zur Deckung zu bringen oder wenigstens partiell zu verknüpfen.

Denn es gibt eine ganze Reihe von Merkmalen, die das psychische Phänomen „Bewußtsein“ so rätselhaft machen – wobei ich zunächst nur das wichtigste dieser Merkmale nennen möchte: seine *Homogenität*. Phänomenologisch gesehen ist Bewußtheit eine bruchlose Qualität, die unseren subjektiven Erlebnisraum durchzieht und ihn zu einem *integrierten Feld*[13] macht. Diese phänomenale „Feldqualität“ ist sehr konkret und läßt sich introspektiv deutlich vom intentionalen Gehalt der durch sie verbundenen mentalen Zustände unterscheiden. Sie kann anstrengungslos und diffus über den Erlebnisraum verteilt sein, aber auch bei Orientierungsreaktionen oder absichtlicher Konzentration – durch den Einsatz „subjektiver Energie“ – fokussiert werden. Mit der inneren Erfahrung dieser „Feldqualität“ geht die „Unteilbarkeits-Intuition“ einher, die Descartes in der *Sechsten Meditation* so eindrücklich formuliert und zu einem Unterscheidungskriterium gegenüber allen ausgedehnten Dingen entwickelt hat: *Bewußtsein ist unteilbar.*

> Nun bemerke ich hier erstlich, daß ein großer Unterschied zwischen Geist und Körper insofern vorhanden ist, als der Körper seiner Natur nach stets teilbar, der Geist hingegen durchaus unteilbar ist. Denn, in der Tat, wenn ich diesen betrachte, d. h. mich selbst, insofern ich nur ein denkendes Ding bin, so vermag

[13] Wie man eine durch viele empirische Befunde gestützte integrierte Feldtheorie des Bewußtseins als einer distribuierten Eigenschaft in einem neuronalen Netz entwickeln könnte, zeigt Marcel Kinsbourne. Ein solcher Ansatz hat den Vorteil, ohne Metarepräsentation und die Gefahr infiniter Regresse auszukommen. Wenn Bewußtsein eine Netzwerkeigenschaft ist, die in multimodalen, nicht-restringierten Repräsentationsfeldern auftritt, dann fragt sich allerdings, ob hier mit dem Begriff „Feld“ nicht eine unhaltbare Anleihe bei der Physik gemacht worden ist. Vgl. Kinsbourne 1988.

> ich in mir keine Teile zu unterscheiden, sondern erkenne mich als ein durchaus einheitliches und ganzes Ding. Und wenngleich der ganze Geist mit dem ganzen Körper verbunden zu sein scheint, so erkenne ich doch, daß, wenn man den Fuß oder den Arm oder irgendeinen anderen Teil des Körpers abschneidet, darum nichts vom Geiste weggenommen ist. Auch darf man nicht die Fähigkeiten des Wollens, Empfindens, Erkennens als seine Teile bezeichnen, ist es doch ein und derselbe Geist der will, empfindet und erkennt. Im Gegenteil aber kann ich mir kein körperliches, d. h. ausgedehntes Ding denken, das ich nicht in Gedanken unschwer in Teile teilen und ebendadurch als teilbar erkennen könnte, und das allein würde hinreichen, mich zu lehren, daß der Geist vom Körper gänzlich verschieden ist, wenn ich es nicht anderswoher zur Genüge wüßte.[14]

Diese Grundannahme neuzeitlicher Subjektivitätstheorien ist mittlerweile durch eine Vielfalt von Untersuchungen, zum Beispiel an hypnotisierten Versuchspersonen oder *split-brain*-Patienten[15] falsifiziert worden. Mit ihr wurde auch die altehrwürdige *Selbsttransparenz des Bewußtseins* obsolet (in den Abschnitten 2.3.2 und 3.2.2 werde ich einige der bedrückenden neuropsychologischen Befunde schildern, die jede diese beiden intuitiven Annahmen voraussetzende oder explizierende Theorie des Geistes unhaltbar machen). All das sollte uns jedoch nicht darüber hinwegtäuschen, daß Bewußtsein eines der prägnantesten Charakteristika unseres normalen mentalen Lebens ist. Die rätselhafte Natur des introspektiven Gegebenseins der Gerichtetheit ist ja gerade eine der Hauptquellen unseres Interesses an einer philosophischen Psychologie. Deshalb sollten wir auch nicht vorschnell zu sprachanalytischen Auflösungen des Problems übergehen. Wenn Bewußtsein tatsächlich keinen Platz in einer empirisch fundierten Taxonomie mentaler Zustände finden kann, dann muß uns eine befriedigende naturalistische Theorie des Geistes zumindest erklären können, wie all die Intuitionen zustande kommen konnten, die es uns als ein so zentrales und bedeutsames Element unserer Innerlichkeit erscheinen lassen. Ich werde zu diesem Zweck in Abschnitt 2.1.4 einige Überlegungen zum Problem des Bewußtseins aus der Perspektive einer Theorie mentaler Repräsentation anstellen. Die von anti-mechanistischen Philosophen immer wieder ins Feld geführte Homogenität und Atomizität phänomenalen Bewußtseins begegnet uns wieder auf einer etwas tiefer liegenden phänomenologischen Analyseebene, nämlich dann, wenn wir spezifische qualitative

[14] Descartes 1972 (1641). Ein klare analytische Rekonstruktion von Descartes' Beweis für den metaphysischen Dualismus bietet Beckermann 1986a.

[15] Hierbei handelt es sich um Patienten, denen (häufig wegen einer schweren, therapieresistenten Epilepsie) das die beiden Großhirnhemisphären verbindende Corpus callosum chirurgisch durchtrennt werden mußte. Vgl. z. B. Bogen 1969a, 1969b, Bogen et al. 1966, Gazzaniga/Le Doux 1978, Le Doux/Wilson/Gazzaniga 1977, Sperry 1970, 1974. Auch Thomas Nagel hat die Implikationen der empirischen Resultate für den philosophischen Begriff des Selbstbewußtseins analysiert; vgl. Nagel 1971. Weitere philosophische Kommentare finden sich bei John Eccles in Popper/Eccles 1982: 380ff (bezüglich der Wiederauferstehung der cartesianischen „Zirbeldrüsen-Hypothese" in Form des „Liaison-Hirns" siehe auch 434ff), Metzinger 1985: Kapitel 4, Puccetti 1973, 1981 und Wilkes 1988: Kapitel 5.

Inhalte unseres inneren Raumes zu verstehen versuchen – wenn wir sozusagen vom weißen Lichtstrahl des Projektors zu den Farben auf der Leinwand übergehen.

Denn **drittens** sind die uns interessierenden fraglichen Zustände auch Erlebnisse von Personen. Das heißt zuallererst: In ihrer Beschreibung müssen wir auf das jeweilige physische System *als Ganzes* rekurrieren. Gehirne oder ihre Subsysteme sind keine Erlebnissubjekte, innere Erfahrungen können ihnen nicht zugeschrieben werden. Denn es ist niemals der visuelle Cortex, der ein Farberlebnis hat – kein von *subpersonalen* Beschreibungsebenen[16] erfaßter Gegenstand kann, so scheint es, als logisches Subjekt von Erlebniszuschreibungen fungieren. Das Roterlebnis hat immer die Person als Ganze. Außerdem besitzen – und das ist wichtiger – subjektive Zustände eine spezifische Form von Gehalt, die ich ab jetzt als ihren phänomenalen oder qualitativen Gehalt bezeichnen werde. Es ist die Schmerzhaftigkeit von Schmerzerlebnissen oder die Qualität von Türkis in einem visuellen Wahrnehmungserlebnis, die den Wesenskern, das eigentlich „Erlebnishafte" vieler unserer inneren Zustände ausmachen. Genau jener Kern ist es, den uns eine überzeugende wissenschaftliche Theorie subjektiven Bewußtseins erklären müßte – und es gibt starke Argumente dafür, daß sie dies aus vielleicht *prinzipiellen* Gründen niemals könnte. Im vorletzten Kapitel werden diese anti-naturalistischen Argumente zum Gegenstand der Überlegungen werden. Qualia – mentale Zustände mit qualitativem Erlebnischarakter – bilden die dritte große Facette des philosophisch-wissenschaftlichen Problemkomplexes „Subjektivität". Was ist das „Erlebnishafte" an unseren inneren Erfahrungen, was müßte man eliminieren, um sie zu blossen Informationsverarbeitungsereignissen zu machen, die auch in einem Computer ablaufen können?[17] Warum genau würden wir von einer

[16] Die Inkompatibilität von personalen und subpersonalen Beschreibungen informationsverarbeitender Systeme ist eine der wichtigsten neueren Varianten des Leib-Seele-Problems. Daniel Dennett hat diesen Aspekt der Fragestellung seit 1969 in vielfältiger Weise immer wieder analysiert; vgl. Dennett 1969, 1981, 1987. Diese begriffliche Inkompatibilität liegt der *nomologischen Inkommensurabilität* des Mentalen zugrunde: Wir können keine detaillierten Kausalketten über die Beschreibungsebenen hinweg verfolgen und keine strikten psychophysischen Gesetze formulieren. Vgl. Davidson 1970, Lanz 1989, Metzinger 1985. Die prominentesten Versuche zur Lösung des Problems bestehen in dem Versuch, einen *nicht-reduktiven Materialismus* zu entwickeln, der keine strikten psychophysischen Gesetze mehr annimmt und trotzdem die wesentlichen monistischen Intuitionen rettet. Kanonische Texte hierzu sind Haugeland 1982 und etwa Kim 1978, 1979, 1982, vgl. aber auch Kim 1990.

[17] Maloney argumentiert dafür, daß die intentionale Zustände begleitenden Qualia durch die Syntax (also letztlich physikalische Eigenschaften) der von dem betreffenden Repräsentationssystem eingesetzten *tokens* von „Mentalesisch" – also seines diskreten internen Codes – konstituiert werden. *„To use a mentalese token embodied as a token of a specified physical type is to experience the phenomenological qualia associated with the occurrence of the thought employing that mentalese token. So, given this hypothesis, the syntax, not the semantics, of mentalese terms constitutes the qualia of the thoughts deploying those terms."* (Maloney 1985: 31) Dieser Gedanke weist – auf der Grundlage einer grundverschiedenen Theorie mentaler Repräsentation – Ähnlichkeiten zu der von mir vorgeschlagenen Hypothese bezüglich des Formats mentaler Präsentate auf. Vgl. Abschnitt 2.1.3.

Maschine, die mit einer Kamera ihre Umgebung betrachtet, die von den Oberflächen der Objekte abgestrahlten Wellenlängen des Lichts analysiert und uns dann über ein Sprachmodul mitteilt *„Ich sehe einen roten Apfel“*, niemals annehmen, daß sie ein Roterlebnis in sich erzeugt hat?

Die Antwort lautet: Weil wir Maschinenzustände aufgrund anderer Eigenschaften individuieren, als unsere eigenen subjektiven Zustände.[18] Die Zustände einer Maschine – zum Beispiel eines natürlichsprachigen Expertensystems zur Analyse visueller Umwelten – würden wir anhand ihrer *kausalen* und *funktionalen* Merkmale kategorisieren. Die inneren Zustände einer Maschine wären einem bestimmten Typ von Zustand genau deshalb zugeordnet, weil sie einen bestimmten Output durch das Sprachmodul verursachen und weil sie mit anderen inneren Maschinenzuständen funktional vernetzt sind. Ihre Identität und ihre Zugehörigkeit zu einer bestimmten Klasse von Zuständen („Rotwahrnehmungen“) würde von uns über ihr kausales Profil festgelegt. Eine etwas weitergehende wissenschaftliche Taxonomie interner Maschinenzustände könnte auch noch die computationalen Relationen, in die solche Zustände eintreten können, analysieren und dabei Begriffe wie „Information“ oder „Repräsentation“ zum Einsatz bringen. Also könnten wir in Einklang mit unseren jeweiligen Interessen verschiedene Beschreibungsebenen und begriffliche Instrumentarien auswählen, um die relevanten Aspekte des systeminternen Geschehens zu beschreiben und vorherzusagen.

Die phänomenale Ontologie dagegen scheint *prima facie* starr zu sein: Subjektive Erlebnisse der Röte eines Apfels unterscheiden sich allem Anschein nach dadurch von den Wahrnehmungen derselben dispositionalen Objekteigenschaft (die auch eine Maschine, d. h. ein durch eine funktionale Analyse erschöpfend zu beschreibendes Repräsentationssystem leisten kann), daß sie durch eine weitere, essentielle Eigenschaft charakterisiert werden können. Diese Eigenschaft ist ihr qualitativer oder phänomenaler Gehalt. Wahrnehmungen erzeugen intentionalen Gehalt, sie sind immer Wahrnehmungen von etwas sinnlich Gegebenem. Wahrnehmungserlebnisse entstehen aber dadurch, daß zu diesem ein *qualitativer* Gehalt hinzutritt. Das heißt: Der den intentionalen Gehalt tragende mentale Zustand ist ein *Quale*, er besitzt eine monadische, intrinsische Eigenschaft (*„Röte“*), die zu seinen relationalen, kausalen und funktionalen Eigenschaften hinzutritt. Qualia sind, das wird nun deutlich, genau jene subjektiven Zustände, die dadurch, daß sie Gegenstände eines bis jetzt noch unbekannten Vorgangs werden (den man vielleicht in einer ersten Vermutung als innere Wahrnehmung bezeichnen könnte), zu Trägern von Qualitäten werden.

Mentale Zustände sind in traditioneller Sichtweise intentionale Zustände, sie sind gerichtet auf etwas außerhalb ihrer selbst und in einem sehr speziellen Sinn enthalten sie es. In moderner Terminologie kann man sa-

[18] Eine kurze und klare Darstellung dieses Problems findet sich bei Churchland 1981, 1989.

gen: Als Repräsentate bilden sie für ein bestimmtes System gewisse relevante Eigenschaften ihrer Repräsentanda in Gestalt eines internen Zustandes ab. Mentale Zustände sind aber nicht nur in manchen informationsverarbeitenden Systemen auftretende Darstellungen von Teilen der Welt, sie sind gleichzeitig auch Erlebnisse von Personen. Darum erheben sich die folgenden Fragen: Was ist es, das zu ihrem intentionalen Gehalt hinzutritt und sie zu subjektiven inneren Repräsentaten macht? Wodurch werden die Resultate mentaler Repräsentationsprozesse zu bewußten Erlebnissen? Wir wissen, daß diese beiden für den Philosophen interessantesten Eigenschaften psychischer Zustände – nennen wir sie „Repräsentationalität" und „Phänomenalität" – disjunkt, also nicht notwendig miteinander verknüpft sind. Denn wir kennen viele Fälle, in denen Systeme Repräsentate erzeugen, ohne dabei Erlebnisse zu haben. Das gilt etwa für Taschenrechner oder Geldautomaten. Es gilt aber auch für menschliche Gehirne im Tiefschlaf, die die Aktivität von Leber, Herz und anderen Organen über die Regulation innerer Repräsentationen dieser Aktivität steuern. Menschliche Gehirne erzeugen Subjektivität nicht permanent, sondern in Episoden. Die wichtige Erkenntnis, daß wir nur episodisch als die psychologischen Subjekte des Wachzustandes existieren, ist vielleicht ein starker Hinweis darauf, daß diese evolutionsgeschichtlich neue Eigenschaft nur unter Bedingungen einer maximalen funktionalen Komplexität aufrecherhalten werden kann.

Die Episodizität subjektiver Zustände wird von anti-naturalistischen Theorien des Geistes gerne unterschlagen. Das ist verständlich, denn die Tatsache, daß natürliche Repräsentationssysteme wie Tiere oder Menschen sich nur intermittierend in phänomenale Systeme verwandeln, verweist deutlich auf die biologischen Wurzeln von Selbstbewußtsein. Eine der erstaunlichsten und bisher viel zu wenig untersuchten Leistungen von Gehirnen ist in diesem Zusammenhang das Aufwachen. Gehirne sind nämlich in der Lage, zwischen verschiedenen Bewußtseinszuständen hin- und herzuspringen, wenn dies erforderlich ist. Aufwachen ist deswegen ein interessantes psychisches Phänomen, weil in ihm der Erlebnischarakter mentaler Zustände erzeugt wird: Erwachen ist die Emergenz subjektiven Bewußtseins. Man kann annehmen, daß sich viele niedere Tiere in einem Zustand befinden, der sie zwar zu Repräsentierenden macht, aber nicht zu Erlebenden. Ob es (umgekehrt gedacht) reines Bewußtsein in dem Sinne geben kann, daß manche Systeme – zumindest im Modus logischer Möglichkeit – mentale Zustände besitzen, die überhaupt nicht repräsentieren und trotzdem aperspektivische, gehaltfreie Erlebnisse darstellen, scheint äußerst fraglich. Vielleicht gibt es irgendwo im Universum einen Planeten der Zen-Meister, auf dem die reine Achtsamkeit eines mittelpunktlosen und nicht durch Inhalte zersplitterten Schauens den üblichen Bewußtseinszustand darstellt. Aber nach allem, was wir bisher wissen, muß man sagen: Phänomenalität setzt Repräsentationalität voraus. Subjektive Erlebnisse treten überhaupt nur in solchen Systemen auf, die komplexe innere Darstellungen von Teilen der Welt erzeugen. Damit wird aber auch eine vorläu-

fige Formulierung unseres philosophischen Projekts möglich, in Gestalt der folgenden Fragen: Wie kommt zum intentionalen Gehalt mentaler Zustände der phänomenale Gehalt hinzu? Was sind empirisch plausible und begrifflich überzeugende Möglichkeiten, den enormen Zuwachs an innerem Reichtum in einem informationsverarbeitenden System zu analysieren, den wir üblicherweise als das Auftreten von *Erlebnissen* bezeichnen? Wie werden aus aperspektivischen internen Repräsentationen solche mit *subjektivem Charakter* und *qualitativem Gehalt*? Können wir diese neuen Eigenschaften analysieren durch Kategorien der subpersonalen Beschreibungsebene, gegebenenfalls durch eine Differenzierung der Kategorie der „Repräsentation"?

Phänomenale Qualitäten scheinen das uns eigentlich interessierende Merkmal vieler subjektiver Zustände zu sein. Wenn eine naturalistische Theorie des Geistes uns überzeugen soll, so müßte sie mindestens einsichtig machen können, wie zwei unterschiedliche corticale Erregungsmuster den qualitativen Unterschied zwischen einem Roterlebnis und einem Türkiserlebnis gewissermaßen „von unten" determinieren können. Es scheint nämlich genau dieser Unterschied zu sein, den wir einer Farbenblinden oder einer Maschine niemals begreiflich machen könnten. Was das Türkis einer tropischen Lagune oder die Schmerzhaftigkeit von Zahnschmerzen angeht, sind die meisten von uns natürliche Essentialisten. Man kann uns nämlich alle als natürliche Psychologen verstehen, als Wesen also, die sich ihre eigenen mentalen Zustände immer schon und vor aller theoretischen Reflexion selbst zu erklären versuchen. Als natürliche Psychologen, als lebensweltliche Experten bezüglich unseres Innenlebens geraten wir aber sofort in einen unvermeidlichen Konflikt mit der professionellen Konkurrenz, den *naturalistischen* Psychologen.

Dieser Konflikt entsteht, weil natürliche Psychologen – ganz im Gegensatz zu naturalistischen Psychologen – Introspektion als Erkenntnisinstrument benutzen, um ihre eigenen subjektiven Zustände zu klassifizieren. Das bedeutet, daß natürliche Psychologen neue mentale Zustände auf einer höheren Stufe erzeugen, deren Gegenstände andere mentale Zustände – zum Beispiel Qualia – sind. Ein Materialist würde vielleicht sagen: Sie erzeugen Gehirnzustände, um die Eigenschaften von Gehirnzuständen festzustellen. Weil sich dadurch der repräsentationale Gesamtzustand[19] des Systems verändert, tragen sie bei diesem Unternehmen immer das (natürliche) Risiko, daß es die abzubildenden alten Zustände nun gar nicht mehr gibt. Trotzdem entdeckt das Instrument der Introspektion etwas, das man als „innere natürliche Arten" bezeichnen könnte: Introspektion erzeugt eine *phänomenale Ontologie*, indem sie die auftretenden mentalen Zustände anhand ihres qualitativen Gehalts individuiert[20] und klassifiziert. In

[19] Was „repräsentationale Gesamtzustände" sind, erkläre ich in Abschnitt 2.3.1.

[20] Daß diese phänomenale „Ontologie des Gehirns" nicht durch phänomenologische Fehlschlüsse in unsere theoretische Analyse dieser Zustände übernommen werden darf, betont immer wieder Lycan. Vgl. Lycan 1987, dazu auch Smart 1959.

diesem Fall gilt *esse est percipi*, das Sein und die Erscheinung mentaler Zustände mit qualitativem Gehalt fallen zusammen, weil Erkenntnisinstrument und Erkenntnisobjekt identisch sind. Alle einzelnen Instanzen von Türkiserlebnissen bilden nach der subjektiven Taxonomie eine Klasse, weil sie dasselbe intrinsische, nicht-relationale Merkmal – eben: *Türkisheit* – besitzen.

Dieses nicht-relationale Merkmal, das für den natürlichen Psychologen die fraglichen Zustände eben genau *als* psychische Zustände identifiziert, hat interessanterweise starke Gemeinsamkeiten mit der **zweiten** phänomenologischen und begrifflichen Säule des Subjektivitätsproblems, mit der bereits angesprochenen („einstelligen") Bewußtheit innerer Erlebnisse. **Erstens** erscheint es introspektiv ebenfalls als *homogen*: Jeder Teil eines türkisen Wahrnehmungsobjekts ist auch türkis, jedes Zeitsegment einer Schmerzepisode ist auch schmerzhaft. Eine Quale zeichnet sich dadurch aus, daß ihr individuierendes Charakteristikum bruchlos über die entsprechenden Partitionen des intentionalen Gehalts verteilt ist. **Zweitens** bilden Qualia phänomenale Atome. Dadurch nämlich, daß die Türkis-Qualität oder Schmerzhaftigkeit eines subjektiven Zustandes introspektiv nicht weiter analysierbar – also: in kleinere Erlebniseinheiten und ihre Beziehungen zerlegbar – zu sein scheint[21], entsteht ein Analogon zu der bereits angesprochenen *Unteilbarkeits-Intuition* bezüglich der Bewußtheit unseres inneren Raumes. Qualia erscheinen uns als nicht weiter analysierbare, prinzipiell irreduzible Bestandteile der subjektiven Wirklichkeit. Sie bilden anscheinend die introspektiv unhintergehbaren Bausteine der subjektiven Realität. Und jede Theorie, in der diese Bausteine des subjektiven Erlebnisraumes nicht mehr vorkommen, ist – so scheint es den meisten von uns – a priori *überhaupt* keine Theorie über psychische Zustände mehr. Diese der Unteilbarkeits-Illusion verwandten Intuitionen bezüglich phänomenaler Qualitäten werde ich ab jetzt als *essentialistische* Intuitionen bezeichnen. Solche subjektiven Gewißheiten haben ihre Grundlage in der Tatsache, daß wir unsere eigenen inneren Zustände viel schneller und zuverlässiger über

[21] Es gibt bei näherem Hinsehen eine Reihe von Gegenbeispielen. Unser Farbdiskriminierungsvermögen ist trainierbar, und Akkorde werden zum Beispiel von ungeübten Musikhörern zunächst als homogene, auditorische *Ganzheiten* erlebt. Das trainierte Ohr kann dagegen einzelne Noten unterscheiden, Komponisten sind sogar in der Lage neue, seltene Akkorde zu imaginieren (Vgl. Churchland 1989: 65). Eine naturalistische Qualia-Theorie muß diese Ausnahmen erklären können (vgl. Abschnitt 2.1.3). Ray Jackendoff hat in empirisch-philosophischen Untersuchungen zum Verhältnis von phänomenalem Bewußtsein und Informationsverarbeitung im Gehirn seine These, daß phänomenale Zustände eine *mittlere* Ebene der repräsentationalen Hierarchie bilden, auch für subjektives Musikerleben empirisch belegt. *„Here direct awareness appears to parallel the **musical surface,** the linearly ordered sequence of pitch events (notes and chords): the notes are what we **hear.** The musical surface is not the lowest level of information structure involved in music perception, for, (. . .), a great deal of auditory processing has already taken place to discriminate pitches, pitch simultaneities, attack points, and temporal sequencing. On the other hand, the musical surface is far from the most central level of musical structure, since there are four higher levels of hierarchical musical structure, that are responsible for the understanding of music.*"(Jackendoff 1987: 292; vgl. auch Kapitel 11)

ihre qualitativen Merkmale identifizieren, als über ihr relationales, kausales oder funktionales Profil.

Was kann nun der *naturalistische* Psychologe all dem erwidern? Die Arbeit derjenigen Forscher, die als empirische Psychologen nach einer naturwissenschaftlichen Methodologie Beiträge zu einem tieferen Verständnis mentaler Zustände liefern wollen, besteht zu einem großen Teil in der Erzeugung und dem Testen[22] von Hypothesen über kausale Rollen und funktionale Interdependenzen. *Theoretische* Repräsentation erzeugt eine andere – revidierbare – Ontologie als *mentale* Repräsentation; die empirische Psychologie benutzt einen anderen epistemischen Zugang zu den fraglichen Phänomenen als der introspizierende Alltagspsychologe. Dieser theoretische Zugang wird niemals dieselbe Innerlichkeit erzeugen können wie die systeminterne Selbstwahrnehmung, denn im Gegenteil ist das epistemische Ziel wissenschaftlicher Theorienkonstruktion ja gerade Intersubjektivität. Bedeutet dies nicht, daß die qualitativen Gehalte subjektiven Bewußtseins sich dem naturalisierenden Zugriff der Neuro- und Kognitionswissenschaften prinzipiell entziehen müssen?

Was an diesem Punkt entschieden werden muß, ist die Frage nach der *erkenntnistheoretischen Autorität* des Subjekts bezüglich seiner eigenen Zustände. Muß eine naturalistische Theorie des Geistes die subjektive Innerlichkeit nur über ihre objektiven Entstehungsbedingungen erklären können oder verfehlt sie systematisch eine relevante Klasse von *Tatsachen*, die nur aus der Perspektive der ersten Person epistemisch erschlossen werden können? Berühmte zeitgenössische Anti-Naturalisten wie Thomas Nagel oder Frank Jackson haben dieses Problem in origineller Weise reformuliert und zu beantworten versucht, deswegen werden wir uns ihren Analysen später zuzuwenden haben. Mit der Frage, ob das Subjekt ein bleibendes Loch im wissenschaftlichen Weltbild ist oder nur ein vorübergehender weißer Fleck auf der Landkarte, sind wir am Ende unseres ersten Ausflugs in die Problemlandschaft bereits auf die Tatsache gestoßen, daß Subjektivität nicht nur ein psychisches, sondern auch ein epistemisches Phänomen ist. Damit wird Subjektivität nicht nur zum zentralen Thema einer empirisch orientierten Philosophie der Psychologie, sondern auch direkt relevant für das Projekt einer naturalistischen Erkenntnistheorie. In Entsprechung zu den eben untersuchten drei phänomenologisch fundierten Hauptaspekten des Subjektivitätsproblems muß man nämlich immer auch drei erkenntnistheoretische Fragen aufwerfen. Sie lauten: *Weiß ich um die „Meinigkeit" der Zustände meines Selbst, und zwar in einem philosophisch interessanten Sinn von „Wissen"? Ist „Bewußtheit" ein epistemisches Phänomen; weiß das psychologische Subjekt – in einem philosophisch interessanten Sinn von „Wissen" – um seine eigene Bewußtheit und deren Inhalte? Stellt der quali-*

[22] Es gibt allerdings bereits Versuche, sogar eine *Wissenschaftstheorie* nach dem Muster subsententialistischer, parallel-distribuierter Informationsverarbeitung zu konstruieren, mithin die Arbeitsweise des Gehirns auch im Bereich intersubjektiven Wissenserwerbs zu etablieren. Vgl. Churchland 1989 Teil 2, besonders Kap. 9 – 11.

tative Gehalt phänomenaler Zustände eine Form von Wissen dar, trägt er Information?

Diese erkenntnistheoretischen Fragen stehen nicht im Zentrum der vorliegenden Arbeit, werden uns jedoch häufig aufs neue begegnen. Ich hoffe, daß dieser erste Erkundungsausflug in die Problemlandschaft einen Eindruck von ihrer komplexen Struktur vermittelt hat. Sie bietet reichlich Hinterhalte für metaphysische Guerilleros, und unser Vertrauen in die klassischen geographischen Beschreibungen dieser Landschaft ist neuerdings durch die empirischen Wissenschaften stark erschüttert worden. Wir sind uns andererseits hin und wieder nicht ganz sicher, ob sie überhaupt *in dieser Welt* ist. Auf dem durch die Problemlandschaft führenden Weg zu einer neuen Theorie des Geistes müssen wir Brücken bauen von der Außenwelt in die Innenwelt. Und es hat oft den Anschein, daß uns all diese Brücken niemals dahin führen können, wo wir doch immer schon *sind.* Die trügerische Selbstgewißheit der Introspektion ist uns allerdings unter dem Ansturm neuen, empirischen Wissens über die im Rücken des Erlebnissubjekts agierende neuronale Maschinerie endgültig zerbrochen. Wir dürfen uns jetzt nicht durch unsere eigenen cartesianischen Intuitionen blenden lassen, sondern müssen versuchen, einen neuen und schärferen Blick auf die Konturen der inneren Natur zu erkämpfen, die auf so rätselhafte Weise *unsere* innere Natur ist. Dabei bleibt uns nur die Flucht nach vorne, ein vorurteilsfreies tieferes Eindringen in die Landschaft der theoretischen Probleme der Subjektivität.

1.3 Kriterien für eine Theorie der Subjektivität: *Die Fragen*

Ich werde nun versuchen, aus der vorangegangenen knappen Diskussion der verschiedenen Aspekte des Problemkomplexes „Subjektivität" eine Reihe von Fragen zu extrahieren, die eine zukünftige Theorie des Geistes zu beantworten haben wird. Diese Liste von Fragen wird sicher nicht vollständig sein, vielleicht kann sie uns jedoch in Gestalt eines *Minimalkatalogs* von Kriterien dienen, anhand derer wir Lösungsversuche beurteilen und verbessern können. Diesen Katalog selbst werde ich gliedern in den **logisch-semantischen**, den **erkenntnistheoretischen** und den **psychologischen** Themenbereich. Das Interesse dieser Arbeit gilt primär psychologischer Subjektivität, deshalb werden es hauptsächlich diese Fragen sein, die ich am Ende zu beantworten versuchen werde.

a) Die logisch-semantische Dimension

Zuerst kann man sich fragen, ob das Subjekt ein nicht weiter zu analysierendes logisches Primitivum ist, das *a priori* zum „Mobiliar der Welt" gehört, oder ob wir es hier mit einem Begriff zu tun haben, der im Verlauf des wissenschaftlichen Fortschritts potentiell auch eliminiert werden könnte.

LS 1: *Ist das Subjekt logisch primitiv?*

Weiterhin bedarf die Semantik des indexikalischen Ausdrucks „Ich" einer Klärung. Wir benötigen ein besseres Verständnis von Sätzen eines bestimmten Typs, nämlich Sätzen in denen das Wort „Ich" auftritt, und zwar bei Selbstzuschreibungen psychologischer Eigenschaften (*„Ich habe gerade Schmerzen"*).

LS 2: *Was sind die Wahrheitsbedingungen solcher Sätze?*

LS 3: *Läßt die Elimination des Subjektgebrauchs von „Ich" eine Lücke in unserem Verständnis offen?*

LS 4: *Kann die Individuierung mentaler Zustände über intrinsische, nicht-relationale Eigenschaften wie ihren qualitativen Gehalt im Rahmen einer wissenschaftlichen Psychologie aufgegeben werden?*

Die letztgenannte Frage betrifft natürlich ebenfalls die psychologische Dimension des Problemraums. Sie ist aber hier als Frage nach der Bedeutung von Termen in einer Wissenschaftssprache gemeint: Angenommen, die Ausdrücke einer psychologischen Theorie beziehen sich nicht mehr auf qualitative Gehalte, in welchem Sinne spricht eine solche Theorie dann *überhaupt* noch von psychischen Zuständen?

LS 5: *Funktionieren mentalistische Terme wie Eigennamen, d. h. spezifizieren sie ihre Referenten über alle logischen Welten hinweg?*

Diese Frage entsteht in Anschluß an Überlegungen von Saul Kripke. Ebenso wie die Thesen von Thomas Nagel versuchen diese Überlegungen eine kohärente Explikation cartesianischer Intuitionen zu liefern. Ich werde die Nagel-Thesen im vierten Kapitel einer gesonderten Betrachtung unterziehen, nachdem die theoretischen Mittel zu ihrer Widerlegung entwickelt sind. Dies tue ich auch deshalb, weil sich Nagels Arbeiten zur Subjektivität mentaler Zustände hervorragend zu einem tieferen Eindringen in die Problematik eignen und ein neues Licht auf bekannte systematische Probleme werfen. Deswegen will ich auch hier schon die Fragen nennen, die direkt auf seine Bemühungen zurückgehen und die am Ende beantwortet werden sollen. Wir brauchen eine

Theorie bestimmter Fälle von *monologisierender Selbstreferenz*, speziell eines bestimmten Typs von *Identitätsaussagen*:

> **LS 6**: *Sind Sätze des Typs „Ich bin TM" ausschließlich als Identitätsaussagen zu interpretieren?*

> **LS 7**: *Gibt es nicht-triviale Wahrheiten in Gestalt solcher Aussagen, die über die reine Selbstreferenz qua historische Person in einem sozialen Kontext hinausgehen? Was wären die Bedingungen der Wahrheit für solche Identifikationen?*

Die zuletzt genannten Fragen sind bereits deutlich erkennbar als Fragen mit einem starken epistemologischen Anteil, sie rücken das Problem, ob es ein *Wissen* bezüglich der uns hier interessierenden psychologischen Eigenschaft geben kann, in den Brennpunkt der Aufmerksamkeit.

b) Die erkenntnistheoretische Dimension

Es fragt sich, ob unsere subjektiven Erlebnisse als solche einen Erkenntnischarakter tragen.

> **E 1**: *Ist „Subjektivität" eine epistemische Relation? Können mentale Repräsentate – ähnlich wie Sätze – als Wahrheitswertträger fungieren?*

> **E 2**: *Was ist der erkenntnistheoretische Status des psychologischen Subjekts? Impliziert die Inkorrigibilität von Selbstzuschreibungen psychologischer Eigenschaften ihre Infallibilität?*

> **E 3**: *Gibt es irreduzible Tatsachen bezüglich der Subjektivität mentaler Zustände, die nur in der ersten Person Singular ausgesprochen und nur aus der Perspektive der ersten Person erkannt werden können?*

Aus dem scheinbar privilegierten Status des psychologischen Subjekts ergeben sich weiterhin Fragen bezüglich einer möglichen *epistemischen Geschlossenheit* des inneren Raumes. Einige von ihnen sind die folgenden:

> **E 4**: *Was bedeutet es, daß subjektive mentale Zustände „private" Zustände sind?*

E 5: *Stellt der qualitative Gehalt mentaler Zustände eine Form von „Wissen über die Welt" dar? Sind Qualia Informationsträger, gibt es so etwas wie „phänomenale Information"?*

E 6: *Ist in der „Meinigkeit", d. h. der erlebnismäßigen Subjektzentriertheit mentaler Zustände eine Form von Wissen kodiert?*

Mit den erkenntnistheoretischen sind schließlich wissenschaftstheoretische Fragen verbunden, die die methodologischen Möglichkeiten und das deskriptive Spektrum einer empirischen Psychologie *überhaupt* betreffen.

E 7: *Läßt sich aus der Subjektivität mentaler Zustände die These ableiten, daß das wissenschaftliche Weltbild prinzipiell unvollständig bleiben muß?*

E 8: *Können introspektive Berichte über innere Eigenschaften in Konkurrenz treten mit Aussagen einer wissenschaftlichen Psychologie?*

c) Die psychologische Dimension

Hier geht es darum, sowohl Eigenschaften mentaler Zustände besser zu verstehen, als auch – wo irgend möglich – die Genese von Intuitionen näher zu beleuchten. Philosophische Probleme kann man begriffsanalytisch kurieren oder in differenziertere Versionen zu transformieren versuchen. Man sollte aber (besonders im Bereich der Philosophie des Geistes) auch versuchen, ihre *introspektiven Wurzeln* freizulegen. Wir alle sind nämlich ständig in Versuchung, in schwierigen Situationen bei der theoretischen Erkundung unserer durch den Problemkomplex „Subjektivität" gebildeten Landschaft einfach „hinzuschauen" und laufen deshalb Gefahr, phänomenologische Fehlschlüsse zu produzieren.[23] Deswegen kann es auch wichtig

[23] Materialistische Philosophen versuchen häufig, die *Prozessualität* des Mentalen ins Blickfeld zu bringen, um einen gewissen Typ von Fehlschluß zu vermeiden. Dieser Fehlschluß besteht darin, von der introspektiven Gewißheit objekthafter Elemente im Bewußtseinstrom auf die Existenz nicht-physischer Gegenstände oder von phänomenalen Eigenschaften auf *reale* zu schließen. Wenn wir ein grünes Nachbild beschreiben, sprechen wir nicht von einem metaphysischen grünen Objekt, sondern von einem „virtuellen" inneren Repräsentationsvorgang, der anderen solcher Vorgänge in Anwesenheit eines bestimmten Typs von externem Wahrnehmungsobjekt ähnelt. Ullian Place hat zu Beginn der Entwicklung der Identitätstheorie auf diesen Typ von *phenomenological fallacy* hingewiesen (vgl. Place 1956). Ein Beispiel für eine neuere Argumentation gegen die Existenz von phänomenalen Individuen findet sich bei Lycan 1987, Kapitel 8.

sein, die *genetischen* Bedingungen introspektiver Gewißheiten näher zu untersuchen.

> **P 1**: *Wie entstehen cartesianische Intuitionen wie z. B. die Kontingenz-Intuition, die Unteilbarkeits-Intuition, das Gefühl der direkten Gegebenheit mentaler Inhalte?*

Weiter muß man fragen, welche neuen psychologischen Eigenschaften durch selbstreferentielle Sprechakte eines gewissen Typs ins Spiel kommen.

> **P 2**: *Welche psychologischen Eigenschaften entstehen dadurch, daß eine Person mental oder extern in Form monologischer Selbstreferenz auf sich Bezug nimmt?*

Dann stellt sich die Frage nach möglichen Erklärungen für die jeweiligen Aspekte des psychischen Phänomens Subjektivität im Rahmen einer naturalistischen Theorie des Geistes.

> **P 3**: *Was ist der „phänomenale Gehalt" mentaler Zustände im Gegensatz zu ihrem „intentionalen Gehalt"? Gibt es mentale Zustände ohne phänomenalen Gehalt?*

> **P 4**: *Koinzidiert das Auftreten von phänomenalem Gehalt immer mit dem, was wir „Bewußtsein" nennen? Gibt es Gegenbeispiele?*

> **P 5**: *Ist das Auftreten von Qualia an eine subjektive Erlebnisperspektive oder an bestimmte Klassen von Bewußtseinszuständen gebunden? Kann es nicht-subjektzentrierte Qualia geben?*

Es fragt sich zudem, welchen Status der Begriff „Bewußtsein" in einer naturalistischen Theorie des Geistes einnehmen kann. „Bewußtsein" ist ein phänomenologisch-alltagspsychologischer Terminus, der semantisch unscharf ist und nur schwer auf eine empirisch fundierte Taxonomie mentaler Zustände abgebildet werden kann.[24] Es ist aber ein Begriff, der in einer philosophischen Theorie des Geistes einen zentralen Platz einnehmen wird, weil er eine wichtige Rolle in unserem lebensweltlichen Selbstverständnis spielt und deshalb – auch wenn ihn eine repräsentationale Theorie der Subjektivität letztlich eliminieren sollte – allein unter anthropologischer Hinsicht bedeutsam ist. Der begriffliche Kommentar zum Fortschritt

[24] Vgl. Allport 1988, Bieri 1991, Wilkes 1984, 1988 und Abschnitt 2.1.4.

der Neuro- und Kognitionswissenschaften muß, wenn die ersten empirisch gehaltvollen Theorien über das Entstehen von Bewußtsein beim Menschen in Sicht kommen, trotz aller prinzipiellen Schwierigkeiten immer wieder versuchen, begriffliche Brückenschläge zwischen einer mit empirischen Erkenntnissen kompatiblen *Representational Theory of Mind* und der sich parallel zu ihr entwickelnden Anthropologie, der Metatheorie des Menschen als eines bewußten Wesens, zu leisten. Dies kann uns allerdings zu begrifflichen Differenzierungen auch auf der Ebene nicht-wissenschaftlicher Diskurse zwingen.

P 6: *Kann es Bewußtsein ohne qualitativen Gehalt geben?*

P 7: *Kann es aperspektivisches Bewußtsein geben?*

P 8: *Wie entsteht ein zentriertes Bewußtsein? Wie ist in bestimmten informationsverarbeitenden Systemen die Emergenz eines phänomenalen Standpunkts bzw. einer Erlebnisperspektive möglich?*

P 9: *Wodurch werden manche Resultate interner Informationsverarbeitung zu* ***meinen*** *Zuständen? Wie entsteht die „Meinigkeit" mentaler Zustände?*

P 10: *Koinzidiert die fragliche Relation mit dem Auftreten von „Bewußtsein"; kann es unbewußte „Meinigkeit" geben?*

P 11: *Ist die Subjektzentriertheit mentaler Zustände eine mit ihrem Erlebnischarakter notwendig verknüpfte Eigenschaft oder kann es Subjektivität ohne qualitativen Gehalt geben?*

Wenn uns eine naturalistische Psychologie das Wesen der verschiedenen Aspekte menschlichen Selbstbewußtseins und ihre wechselseitigen Beziehungen erläutern kann, stellt sich auch die für den philosophischen Geistbegriff bedeutsame Frage nach einer universellen Psychologie: Können wir *mentale Universalien* über spezies- und hardware-unabhängige Eigenschaften definieren?

P 12: *Welche Klasse von Systemen kann die unter dem Begriff „Subjektivität" gebündelten Eigenschaften instantiieren? Ist künstliche Subjektivität möglich?*

Die eben aufgelisteten Fragen sollen einen Minimalkanon der systematischen Probleme darstellen, die eine philosophische Theorie des Erlebnissubjekts lösen muß. Natürlich decken sie nur einen kleinen Bruchteil des Problemfelds ab. Sie können uns aber als inhaltliche Richtschnur bei den nun folgenden Untersuchungen dienen und als roter Faden, der am Ende wieder aufgenommen wird.

Es werden am Ende jedoch *zwei* rote Fäden sein, die wir wieder aufnehmen und miteinander verbinden müssen: Zusätzlich zu den eben aufgeworfenen systematischen Fragen werde ich im vierten Kapitel die Argumente des in der Gegenwart vielleicht bedeutendsten anti-naturalistischen Philosophen des Geistes näher untersuchen. Auf diese Weise hoffe ich, ein zweites Sprungbrett in die Diskussion zu entdecken und durch Einbeziehung der philosophischen Fachdiskussion zusätzliche Perspektiven auf die Probleme zu gewinnen, die den Aufbau einer Theorie des Erlebnissubjekts als so dringend notwendig erscheinen lassen. Es gibt deshalb in diesem Buch zwei Gruppen von Fragen und zwei Gruppen von Antworten: Allgemeinsystematische Fragen und Antworten sowie solche, die sich speziell mit Thomas Nagels Versuchen auseinandersetzen, die Dimension des Subjektiven gegen vorschnelle Reduktionsversuche zu verteidigen. Aus diesem Grund nenne ich der Vollständigkeit halber bereits hier die drei zentralen Fragen der zweiten Gruppe, obwohl sie erst im vierten Kapitel weiter entfaltet und untersucht werden. Leser, die mit Nagels Schriften nicht vertraut sind, sollten sich dadurch nicht beunruhigen lassen. Die Bedeutung von Nagels Beiträgen wird im vierten Kapitel deutlich werden.

P 13: *Gibt es für externe Beobachter eine zuverlässige Möglichkeit, um zu entscheiden, ob in einem gegebenen System mit bekannten physischen und funktionalen Eigenschaften die phänomenale Qualität des Nagelschen „Wie es ist, ein X zu sein" auftritt?*

P 14: *Lassen sich das Nagelsche „objective self" und die mit seinem Auftreten verbundenen Perspektivenwechsel als Resultat mentaler Repräsentation verstehen?*

LS 8: *Was drückt der „philosophische Gedanke" aus? Was heißt: „Dieses denkende Subjekt betrachtet die Welt durch die Person TN"? Kann man sinnvollerweise davon sprechen, daß das Subjekt die öffentliche Person als Erkenntnisinstrument einsetzt?*

Thomas Nagel hat (1974) vorgeschlagen, daß wir uns um die Entwicklung einer *objektiven Phänomenologie* bemühen sollten. Eine solche Phänomenologie müßte es uns ermöglichen, den subjektiven Charakter mentaler Zustände weiter zu erforschen ohne dabei auf philosophisch-introspektive

Phantasien oder empathische Imagination zurückgreifen zu müssen.[25] Ob es jemals möglich sein wird, vermittels einer objektiv-theoretischen Terminologie einem andersartigen Wesen subjektive Erlebnisqualitäten begreiflich zu machen, die von seinem Nervensystem niemals erzeugt werden können, scheint in der Tat recht fraglich. Nagel hat jedoch mit Sicherheit recht, wenn er darauf hinweist, daß das Potential an Präzision und Beschreibungsfähigkeit unserer objektiven Begrifflichkeit noch lange nicht ausgeschöpft ist.[26] Dabei könnte eine Analyse von *strukturellen* Eigenschaften[27] der phänomenalen Realität erfolgversprechender sein als die direkte Behandlung subjektiver Qualitäten. Dies ist ein Gedanke, der, wie in den folgenden Kapiteln deutlich werden wird, meine eigenen Überlegungen sehr stark bestimmt hat.

In Zusammenhang mit dem *self-locating thought*, der das objektive Selbst mit der partikularen Perspektive z. B. von TN verbindet, sagt Nagel (1986), daß dieser Gedanke bzw. die mit ihm verknüpfte mentale Operation etwas Wesentliches über uns enthüllt. Er enthüllt eine *Kompetenz für Objektivität* – auch uns selbst gegenüber – die, so Nagel, allem Anschein nach nicht auf einfachere und weniger kreative mentale Vorgänge zurückgeführt werden kann.[28] Wenn aber der von Nagel umkreiste dunkle Aspekt unseres Innenlebens überhaupt korrekt als eine „mentale Operation" höchster Stufe beschrieben werden kann, dann muß eine am Modell der Informationsverarbeitung orientierte Theorie des Geistes in der Lage sein, die erlebnismäßige Atomizität dieses Aspekts plausibel aufzulösen in Relationen zwischen theoretischen Elementen einer niedrigeren Beschreibungsebene. Mit anderen Worten: Wenn die menschliche Fähigkeit, auf sich selbst immer wieder

[25] Diesbezüglich hat Janet Levin argumentiert, daß eine sehr grosse Anzahl von ganz unterschiedlichen subjektiven Erfahrungen den Aufbau des begrifflichen „Rohmaterials" für *theoretisches* Wissen (etwa bezüglich der subjektiven Erlebnisse von Fledermäusen) ermöglicht, ein Wissen, das die Einordnung von beliebigen Erlebnissen entlang gewisser Dimensionen wie Kompatibilität, Intensität oder funktionalem Profil ermöglicht. Levin sagt, daß die entscheidende Äquivokation des Nagelschen und Jacksonschen Begriffs von *„innerem Wissen"* in dem Besitz des Begriffs eines mentalen Zustands und der Fähigkeit, ihn auf eigene mentale Zustände erfolgreich *anzuwenden* besteht. Deswegen ist sie, auch mit Blick auf aus der menschlichen Perspektive als phänomenale Exoten erscheinende Systeme, optimistisch bezüglich des von Nagel angesprochen Projekts der objektiven Phänomenologie: *„My suspicion, however, is that these more exotic cases seem more intractable because we know quite little about the relevant dimensions of alien experiences, and even less about how these experiences are to be ordered along those dimensions. Our current lack of knowledge may indeed be due to a gap in our objective theories, but there is no reason to think that it cannot be overthrown by acquiring more information of a perfectly objective sort."* Vgl. Levin 1990 (1986): 486.

[26] Häufig haben Vetreter des eliminativen Materialismus darauf hingewiesen, daß die Fähigkeit zur genauen Beschreibung unseres Innenlebens nach einer „reduktionistischen Revolution" – gefolgt von der Elimination des cartesianisch geprägten Alltagsidioms – möglicherweise wesentlich gesteigert würde, weil wir mit der viel differenzierteren *neurowissenschaftlichen* Taxonomie einen wesentlich exakteren Begriffsapparat zur Beschreibung innerer Erlebnisse zur Verfügung hätten. Vgl. z. B. Churchland 1989: 53ff.

[27] Vgl. Nagel 1974, in Bieri 1981: 272.

[28] Vgl. Nagel 1986: 66 (1992: 117).

eine quasi-objektive Perspektive einzunehmen, *überhaupt* eine natürliche Eigenschaft unserer Spezies ist, dann sollte es auch prinzipiell möglich sein, eine natürliche Erklärung für dieses hochinteressante Phänomen zu finden. In diesem Sinne werde ich nun die Bestandsaufnahme in bezug auf unser Problem abschließen und in den nächsten beiden Kapiteln untersuchen, ob im Rahmen einer empirisch plausiblen Theorie mentaler Repräsentation Einsichten über Struktur und Entstehungsbedingungen des inneren Erlebnisraums gewonnen werden können, die neue Antworten auf die eben formulierten Fragen zulassen. Einige solcher Antworten werde ich dann im Schlußkapitel vorstellen.

2. Kapitel

Mentale Repräsentation und phänomenale Zustände

Auf den folgenden Seiten werde ich das Problem der Subjektivität geistiger Zustände aus der Perspektive einer naturalistischen Theorie mentaler Repräsentation betrachten. Zunächst werde ich einige allgemeine Überlegungen zur Frage der internen Repräsentation von Teilen der Welt durch mentale Zustände anstellen. Diese Überlegungen werden dann zu einer Rekonstruktion von mentaler Repräsentation als dem Sonderfall eines umfassenderen Vorgangs – nämlich: *mentaler Simulation* – führen. Da ich in dieser Arbeit primär an dem *relationalen* Aspekt von Subjektivität interessiert bin, treten die Problemfelder *Bewußtsein* und *Qualia* dabei eher in den Hintergrund. Trotzdem werde ich in zwei kürzeren Abschnitten (2.1.3 und 2.1.4) Andeutungen darüber machen, wie man diese Probleme im Rahmen einer Theorie mentaler Repräsentation lösen könnte. Schließlich möchte ich über den in diesem zweiten Kapitel eingeführten Begriff des „mentalen Modells" versuchen, zu einem besseren Verständnis der Frage zu gelangen, was es heißt, daß manche informationsverarbeitenden Systeme – wie zum Beispiel menschliche Organismen – in der Lage sind, komplexe innere Repräsentate ihrer *selbst* zu erzeugen. Vor dem Hintergrund der im dritten Kapitel aufgestellten These, daß mentale Selbstmodellierung der Schlüssel zum Verständnis des psychischen Phänomens „Subjektivität mentaler Zustände" ist, werde ich dann abschließend versuchen, Antworten auf die im ersten Teil formulierten philosophischen Fragen zu geben.

2.1.1 Mentale Repräsentation: Intentionalität durch Informationsverarbeitung

Mentale Repräsentation ist ein Vorgang, durch den manche Biosysteme innere Beschreibungen von Teilbereichen der Wirklichkeit erzeugen.[1] Sie erzeugen diese Beschreibungen als mentale Zustände, also in Form möglicher *Bewußtseinsinhalte*. Die Inhalte unseres Bewußtseins sind die Resultate einer von unseren Gehirnen erbrachten Repräsentationsleistung, deren Mechanismus wir derzeit immer besser zu verstehen beginnen. Die Erzeugung mentaler Zustände ihrerseits ist aber nur ein Sonderfall biologischer

[1] „Beschreibung" ist hier in einem weiten Sinne gebraucht und meint nicht die Erzeugung linguistischer bzw. propositionaler Repräsentate. Wie in den Abschnitten 2.2 und 2.2.1 deutlich wird, unterscheiden sich die durch den Prozeß der mentalen Repräsentation erzeugten inneren Strukturen von Beschreibungen in Form von Sätzen (etwa einer *lingua mentis* oder einer Fodorschen *language of thought*) gerade dadurch, daß sie nicht auf Wahrheit abzielen, sondern auf *Ähnlichkeit*.

Informationsverarbeitung: Der überwiegende Teil der Repräsentation von Eigenschaften der Welt durch das Hervorbringen spezifischer innerer Zustände geschieht in völliger Abwesenheit von phänomenalen Qualitäten und subjektivem Bewußtsein. Die komplizierten Vorgänge interner Informationsverarbeitung, die etwa notwendig sind, um unseren Herzschlag zu steuern oder die Aktivität des Immunsystems[2] zu regulieren, erzeugen eine Vielzahl von internen Systemzuständen, welche niemals zu Inhalten subjektiven, phänomenalen Bewußtseins werden.

Die erste Frage in Zusammenhang mit dem Phänomen der mentalen Repräsentation lautet also: Was macht *interne* Repräsentationen zu *mentalen* Repräsentationen, zu Prozessen, die *potentielle Inhalte subjektiven Bewußtseins* sind? Daß biologische Nervensysteme interne Abbildungen der Welt und ihrer kausalen Matrix in Gestalt von Zuständen mit Stellvertreterfunktionen generieren können, ist eine Tatsache, die ich hier nicht weiter problematisieren werde. Das erstaunliche Faktum dagegen, daß solche internen Abbildungen von Teilen der Welt – zumindest in unserem eigenen Fall – zu den *Erlebnissen von Personen* werden können, lenkt unsere Aufmerksamkeit auf einen der zentralen Aspekte der Subjektivitätsfrage, nämlich die Inkompatibilität der personalen und der subpersonalen Beschreibungsebene.[3] Dieser Aspekt konfrontiert uns mit einer weiteren Variante

[2] Nicht alle biologische Informationsverarbeitung ist neuronale Informationsverarbeitung. Es gibt nicht-lokalisierte Informationsträger wie zum Beispiel die Hormone, und es gibt auch nicht-lokalisierte – aber ebenfalls steuerungsbedürftige – *Organe*, wie das Immunsystem. Dies ist, nebenbei bemerkt, eines der wichtigsten Merkmale, die die bis heute diskutierten neuronalen Netze vorerst zu biologisch eher unrealistischen theoretischen Modellen machen. Außerdem ist die *funktionale Architektur* des Gehirns noch in weiten Teilen eine *terra incognita*, die aus seiner anatomischen Struktur hypothetisch erschlossen wird. Dazu kommt, daß das menschliche Gehirn zu mehr als 50 % aus Gliazellen besteht, deren funktionale Rolle derzeit noch weitgehend unverstanden ist und schließlich leitet es Information auf andere Weise weiter als heutige konnektionistische Systeme. *„Neben diesen (. . .) Unterschieden im Aufbau der Netzwerke unterscheidet sich die Signalübertragung zwischen Recheneinheiten bzw. Neuronen in vielen Punkten: während in konnektionistischen Systemen zum Beispiel von einer stetigen Weitergabe der Aktivierungswerte ausgegangen wird, übertragen Neuronen ihre Erregung in diskreten Impulsen, d. h., die Signale werden in Axonen frequenzmoduliert, die Signale in konnektionistischen Systemen dagegen amplitudenmoduliert übertragen. Ein weiterer Punkt betrifft die Integration der auf eine Recheneinheit bzw. Nervenzelle einwirkenden Signale: In konnektionistischen Systemen werden die eingehenden Signale meist einfach addiert (selten wird eine multiplikative Verknüpfung benutzt). Die Integration der auf eine Nervenzelle einwirkenden Signale ist dagegen weit komplizierter und es ist vollkommen unklar, mit welcher mathematischen Funktion sich diese beschreiben läßt.“* (Helm 1990: 138)

[3] Es ist ein Verdienst Daniel Dennetts, in seinen Analysen diesen Punkt so deutlich hervorgehoben zu haben. Vgl. zum Beispiel Dennett 1969: 93ff, Dennett 1978: 267ff, Dennett 1987: 57ff. Daß wir von unterschiedlichen logischen Subjekten (Personen und subpersonalen Entitäten wie zum Beispiel Gehirnen) prädizieren müssen, ist eines der Hauptprobleme in der modernen Diskussion des Leib-Seele-Problems. Es wurde unter dem Stichwort „nomologische Inkommensurabilität des Mentalen“ durch Autoren wie Donald Davidson und Jaegwon Kim in die Debatte eingeführt und hat zu den verschiedenen Versuchen geführt, einen nicht-reduktiven Materialismus zu entwickeln. Vgl. Davidson 1981 (1970), Horgan 1983, Kim 1978, 1979, 1982b, 1984b, 1985. Zu den persistierenden Schwierigkeiten des Projekts vgl. die *Presidential Address* vor der APA von Kim 1989 und Stephan 1991.

des Leib-Seele-Problems. Diese neue Variante entsteht dadurch, daß wir den subjektiven Charakter mentaler Zustände unter dem Aspekt der Informationsverarbeitung betrachten: In welcher Relation stehen komplexe Informationsverarbeitungsereignisse – zum Beispiel in menschlichen Gehirnen – zu gleichzeitig ablaufenden psychischen Episoden, die von den betreffenden Systemen selbst unter Einsatz *externer* Repräsentationscodes als *ihre* „subjektiven Erlebnisse" bezeichnet werden?

Die Fortschritte der Neuro- und Kognitionswissenschaften haben deutlich gemacht, daß das Auftreten und der Gehalt psychischer Episoden in sehr starker Weise durch Eigenschaften des Informationsflusses im Gehirn determiniert wird. Aus diesem Grunde scheint es sinnvoll, nicht nur mentale Zustände im allgemeinen, sondern auch das unter dem Stichwort „Subjektivität" zusammengefaßte Bündel problematischer Eigenschaften dieser Zustände durch Begriffe derjenigen Beschreibungsebene zu analysieren, die Objekte mit psychologischen Eigenschaften als *informationsverarbeitende* Systeme ins Blickfeld rückt. Die zentrale Kategorie dieser theoretischen Ebene ist das Konzept der *Repräsentation*. Der Begriff der „Repräsentation" hat in unserem Jahrhundert über die Verknüpfung mit dem Begriff der Information eine mathematische Präzisierung und eine empirische Verankerung erfahren, die ihn interessant macht für naturwissenschaftlich orientierte Analysen kognitiver Phänomene und psychischer Zustände im allgemeinen. In der Künstliche-Intelligenz-Forschung, in der Kognitionspsychologie und in verschiedenen Neurowissenschaften spielt der Repräsentationsbegriff heute die zentrale Rolle in der Theoriebildung. Das darf nicht darüber hinwegtäuschen, daß diese Entwicklung zu einer mehr als problematischen semantischen Inflation des Terms geführt hat[4] und daß Information – derjenige Begriff also, der diese Entwicklung in Richtung auf einen Brückenschlag zwischen Natur- und Geisteswissenschaften überhaupt erst möglich gemacht hat – die bei weitem *jüngere* Kategorie darstellt. Repräsentation ist nämlich ein traditioneller *topos* abendländischer Philosophie.[5]

[4] Nützliche Begriffsklärungen und Literaturverweise bezüglich verschiedener Theorien mentaler Repräsentation finden sich in Stich 1992.

[5] Das erste sicher belegbare Auftreten des Begriffs in der abendländischen Geistesgeschichte finden wir bei Cicero, der *repraesentatio* vorwiegend in seinen Briefen und Reden verwendet, weniger in seinen philosophischen Schriften. (Bei diesen und allen folgenden Angaben stütze ich mich auf die detaillierte und äußerst empfehlenswerte Begriffsgeschichte von Eckart Scheerer und Oliver Scholz. Vgl. Scheerer 1990a, 1990b; Scholz 1991.) Einen klar benennbaren griechischen Prototyp des Begriffs *repraesentatio* gibt es nicht, es hat aber den Anschein, als tauchten alle heutigen Bedeutungselemente von „Repräsentation" bereits in der lateinischen Fassung auf. Für die Römer bedeutet *repraesentare* etwas vorher Abwesendes in einem sehr wörtlichen Sinn wieder *in die Gegenwart zu bringen*. Im frühen Mittelalter bezog der Begriff sich überwiegend auf konkrete Dinge und Handlungen; das semantische Element von „Stellvertretung" ist bereits in einem juristischen Text aus dem 4. Jahrhundert belegt (vgl. Podlech 1984: 510). Später hinzutretende Bedeutungselemente sind „Vorstellung" (das sich über den Einsatz von *repraesentatio* für mittelalterliche Mysterienspiele bis in die heutige Parallele zwischen „Theatervorstellung" und „innerer Bühne" *[forum internum]* bzw. die Denkfigur des *Theater Model*

of Mind erhalten hat, gegen die zum Beispiel Lycan 1987 noch polemisiert) und „Darstellung“ (vgl. zum Beispiel die heutige Debatte über *piktoriale* mentale Repräsentation in Block 1981, Finke 1989, Kosslyn 1980, Rehkämper 1990 oder Shephard/Chipman 1970). Wie Scheerer bemerkt, hat sich die *technische* philosophische Bedeutung von „Repräsentation“ erst durch die mittelalterliche Adaption der aristotelischen Philosophie entwickelt, während die terminologischen Wurzeln von „mentaler Repräsentation“ bereits in einer Art „gelehrter Alltagspsychologie“ zu finden sind (vgl. Scheerer 1990a: 8). Der philosophische Term der *repraesentatio* besaß bereits vor der Neuzeit eine enorme Bedeutungsfülle und seine Geschichte bietet einen großen Reichtum an semantischem Wandel. Im Kontext der vorliegenden Arbeit ist die zentrale Bedeutungstransformation vielleicht die Verknüpfung des erkenntnistheoretischen Repräsentationalismus mit dem modernen Bewußtseinsbegriff: Sie ist eine der Wurzeln jener durch Descartes eingeleiteten terminologischen Revolution, die allerdings nicht *nur* eine terminologische, sondern auch eine fundamentale *theoretische* Umwälzung in der metaphysischen Phase der Entwicklung des okzidentalen Geistbegriffs herbeiführte. Die Ideen werden nun in einem nicht-platonischen Sinne als Teile des *individuellen* Bewußtseins einzelner Menschen verstanden – eine gegen die Scholastik gewendete semantische Transformation, die sich bis heute in dem erhält, was von angelsächsischen Philosophen des Geistes als *folk psychology* bezeichnet wird. (Vgl. den klassischen Aufsatz von Churchland 1981a [dazu auch Churchland 1970, 1979, 1984] und die Debatte um den Theoriestatus der Alltagspsychologie in Churchland 1988a und Dennett 1987. Vgl. dazu auch Lycan 1990: Teil VI.) *Représenter* bezeichnet bei Descartes denjenigen Aspekt einer Idee, der mit ihrem *Gehalt* zu tun hat. Für ihn werden Ideen durch ihren repräsentationalen Gehalt individuiert und es kann keine Ideen geben, die *nicht* repräsentieren. Descartes unterscheidet auch bereits deutlich zwischen einem *Akt-* oder *Prozeßaspekt* des Mentalen und seinem *Gegenstands-* oder *Gehaltaspekt*. Wenn Descartes den repräsentationalen Aspekt von Ideen im Auge hat, bezeichnet er sie häufig als „Bilder“ (er spricht sogar bereits von „Bildern im Gehirn“). Scheerer (1990b: 8) zitiert auch eine Stelle aus den Gesprächen mit Burman, in der Descartes Imagination als einen Vorgang beschreibt, bei dem die Seele äußere Gegenstände in ihrer Abwesenheit im Gehirn bildet und repräsentiert. *Représenter* kann also bei Descartes *Verursachung von oben* heißen. Für seinen Kritiker Gassendi sind Ideen dann allerdings wieder ganz und gar *materielle* relationale Strukturen, die durch ihre *eigene* relationale Struktur Objekteigenschaften abbilden und eine streng mechanistisch zu denkende kausale Genese aufweisen – eine deutliche Parallele zum Begriff des *mentalen Modells*, wie er 1943 aller behavioristischen Euphorie zum Trotz von Kenneth Craik wieder eingeführt wurde (vgl. Abschnitt 2.2). Gegen Descartes wies Gassendi darauf hin, daß Ideen selbst auch *teilbar* sein müßten, wenn sie teilbare (körperliche) Dinge repräsentieren sollen. Man kann sagen, daß neben dem Leib-Seele-Problem die Frage nach der Möglichkeit von epistemischen Beziehungen des Egos zu Welt durch *psychophysische Analogien* die größte Schwierigkeit des cartesianischen Substanzdualismus war.

Die begriffliche Trennung der *reprasentatio* von der *similitudo* war in der abendländischen Geistesgeschichte ein langwieriges Projekt. (Bezüglich des weiteren Verlaufs der Begriffsgeschichte von der Leibniz-Rezeption bis zu heutigen Überlegungen betreffs bildhafter Darstellungsformen und der eigentlichen mentalen Repräsentation finden sich reichhaltige Angaben in Scholz 1991. Sie empfehle ich dem historisch interessierten Leser.) Im Lichte neuester empirischer Erkenntnisse über subsymbolische Informationsverarbeitung in neuronalen Netzen und analoge Darstellungsformen mit Hilfe abstrakter Isomorphismen (vgl. etwa Rehkämper 1990, Abschnitte 2.2 und 2.2.1) erscheint die Geschichte des Begriffs „Repräsentation“ bis zum klassischen Idealismus als eine, in der viele wichtige Denkfiguren heutiger materialistischer Debatten in der Philosophie des Geistes bereits auftauchen und in der auf verblüffende Weise Konzeptionen entwickelt worden sind, die wir heute vielleicht sogar als „neurobiologisch realistisch“ einschätzen würden. Erkenntnistheorie und Psychologie waren damals noch keine getrennten Projekte. In einer Zeit, in der sowohl Erkenntnis wie auch das Mentale wieder erfolgreich naturalisiert werden und sich auf diese Weise näher rücken, muten viele der metaphysischen Argumentationsfiguren des Mittelalters und der beginnenden Neuzeit auf seltsame Weise bekannt an.

Um den durch die eben angesprochene semantische Aufblähung verursachten Unschärfen und Vagheiten der aktuellen Debatte zu entgehen, müssen wir zunächst einen Blick auf die *logische Struktur der Repräsentationsbeziehung werfen,* um zu einem konsistenten Begriff der uns interessierenden epistemischen und psychischen Prozesse zu gelangen. Den Begriff der „mentalen Repräsentation" kann man durch eine dreistellige antisymmetrische Relation zwischen Repräsentanda und Repräsentaten bezüglich eines Systems analysieren: Repräsentation ist ein Prozeß, der *für* ein System die innere Beschreibung eines *Repräsentandums* durch die Erzeugung eines als *Repräsentat* fungierenden Zustands leistet.[6] Die „Antisymmetrie" bezieht sich auf die zweite und dritte Argumentstelle der Relation: mit M_{Rep} (S,X,Y) gilt niemals zugleich M_{Rep} (S,Y,X). Im Idealfall bildet eine Teilmenge der Eigenschaften des Repräsentats die relevanten „Zieleigenschaften" ab. Da das Repräsentat ein physischer Teil des jeweiligen Systems ist, verändert sich das System durch den Vorgang der internen Repräsentation ständig selbst: Es erzeugt in sich neue physische Eigenschaften, um Eigenschaften der Welt zu erfassen.

Mentale Repräsentation: $\mathbf{M_{Rep}}$ (S,X,Y)
- X repräsentiert Y *für* S.
- X ist ein interner Systemzustand.
- X ist potentiell introspizierbar; d. h. kann seinerseits zum Repräsentandum von Repräsentationsprozessen höherer Ordnung werden.
- $\mathbf{M_{Rep}}$ ist antisymmetrisch.

Weil im Herzen einer naturalistischen Theorie des Subjekts eine problematische Eigenschaft der Resultate interner Informationsverarbeitung steht, ist es wichtig, Prozeß und Resultat auf analytischer Ebene klar zu trennen. Ein Großteil der gegenwärtigen Diskussion krankt daran, daß nicht deutlich zwischen *Repräsentation* und *Repräsentat* unterschieden wird. Selbstverständlich ist es nicht der neuronale Prozeß *selbst*, der mental ist oder zum Inhalt von Bewußtsein werden kann: Es sind die spezifischen, durch diesen Prozeß erzeugten inneren Aktivierungszustände bzw. Datenstrukturen. Wenn Sie das Buch in Ihren Händen betrachten, sind Sie sich nicht eines hochkomplexen Repräsentationsvorgangs in Ihrem visuellen Cortex bewußt, sondern des *mentalen Modells*[7], welches durch diesen Prozeß generiert wird. Sollten Sie gleichzeitig introspektiv die bei der Lektüre in Ihnen ausgelösten mentalen Zustände – Langeweile, Ablehnung oder Interesse – beobachten, so sind auch dann die Inhalte Ihres Bewußtseins mentale Repräsentate und nicht der jeweilige Konstruktionsprozeß selbst. Subjektivität ist eine Eigenschaft einer bestimmten Klasse von mentalen Repräsentaten.

---

[6] Vgl. Herrmann 1988.

[7] Den Begriff des „mentalen Modells" erläutere ich in Abschnitt 2.2.

Wenden wir uns zunächst kurz denjenigen Relata zu, die zum Gehalt mentaler Repräsentate werden: den Sachverhalten der Welt, die in unserer dreistelligen Relation als *Repräsentanda* fungieren. Bei den Repräsentanda kann es sich um externe Fakten wie die Präsenz eines natürlichen Feindes, einer Nahrungsquelle oder eines Geschlechtspartners handeln, aber auch um Symbole, Argumente oder Theorien über die Subjektivität mentaler Zustände. Interne Fakten wie der Blutzuckerspiegel oder das Vorhandensein von Krankheitserregern können ebenfalls zu Repräsentanda werden, indem sie die Aktivität des zentralen Nervensystems modulieren und so seinen Informationsfluß verändern. Genauso können Relationen wie etwa die Entfernung[8] zu einem bestimmten – ebenfalls intern repräsentierten – Zielzustand und Klassen, zum Beispiel von Lust oder Schmerz produzierenden Verhaltenstypen, zu Objekten des Abbildungsvorganges werden und so höhere kognitive Vorgänge einleiten. Besonders wichtig in Zusammenhang mit dem Problem der Subjektivität ist die Tatsache, daß das System *als Ganzes* mit all seinen internen, öffentlichen und relationalen Eigenschaften zum Repräsentandum werden kann. Repräsentanda können sowohl systemexterne wie auch systeminterne Bereiche der Welt sein.

Bei den *Repräsentaten* handelt es sich dagegen in dem hier betrachteten Sonderfall der mentalen Repräsentation immer um Datenstrukturen, die – ähnlich wie funktionale Zustände – durch interne physikalische Eigenschaften des betreffenden Systems realisiert sind.[9] Die Form der Realisierung kann sehr unterschiedlich sein. Bei von-Neumann-Maschinen kann die betreffende Information in Gestalt diskreter „symbolischer Atome" unter einer bestimmten Adresse abgelegt sein, es ist jedoch auch möglich – und neurobiologisch realistischer – daß ein System distribuierte Repräsentate auf der *subsymbolischen* Ebene erzeugt, zum Beispiel durch die strukturelle Fixierung bestimmter Energieminima und komplexe Konnektivi-

[8] Der theoretische Rahmen des Konnektionismus bietet mathematisch präzise Ähnlichkeitskriterien für die internen Repräsentationen eines Netzwerks an: Wenn man annimmt, daß solche Systeme (etwa ein neuronales Netz) interne Repräsentate als durch Aktivierungsvektoren in n-dimensionalen Vektorräumen beschreibbare Zustände erzeugen, dann kann man die Ähnlichkeit (oder „Entfernung") zwischen zwei Repräsentaten als den *Winkel* zwischen zwei Aktivierungsvektoren analysieren. Für eine philosophische Naturalisierung der Erkenntnistheorie ist diese Tatsache in ihrer Bedeutung kaum zu unterschätzen. Vgl. Abschnitte 2.2 und 2.2.1. Bezüglich der *Identitätskriterien* vgl. Churchland 1998.

[9] Auch bei funktionalen Zuständen, die durch ihre kausale Rolle individuiert werden, spricht man von Realisierung in Form physikalischer Strukturen mit den entsprechenden „kausalen Kräften" und von ihrer *Multirealisierbarkeit*; d. h. von ihrer Eigenschaft, durch eine mehr oder weniger große Anzahl von funktional isomorphen und physikalisch unterschiedlichen Strukturen im kausalen Netzwerk der Welt wirksam werden zu können. Mentale Repräsentate dagegen werden durch ihren Gehalt individuiert. Wenn man davon ausgeht, daß sie als solche auch eine Steuerfunktion innerhalb des gegebenen Systems besitzen, dann können sie dies nur insofern sie als konkrete Strukturen oder Prozesse realisiert sind. Um zu verstehen, wie mentale Repräsentate realisiert werden können, muß man also die *Funktionalität* ihres Gehalts untersuchen. Vgl. Abschnitt 2.2.1.

tätsveränderungen einer Vielzahl einzelner logischer Elemente.[10] Mentale Repräsentate sind also keine nicht-physischen Einzeldinge, sondern interne Zustände physikalischer Informationsverarbeitungssysteme, die von uns unter bestimmten, sehr abstrakten Hinsichten betrachtet werden – nämlich als Träger intentionalen und phänomenalen Gehalts.

Die Repräsentate sind außerdem in der eben angebotenen Darstellung der Repräsentationsbeziehung $\mathbf{M}_{\mathbf{Rep}}$ durch ein *teleologisches Zusatzkriterium* spezifiziert worden. Was soll das heißen? Es bedeutet, daß die von einem System erzeugten mentalen Repräsentate ihren Gehalt *für* das System darstellen: Ein interner Zustand *X* repräsentiert einen Bereich der Welt *Y für* ein System S. Seinen repräsentationalen Gehalt besitzt der betreffende physikalische Zustand innerhalb des Systems also nur im Kontext eines Systems.[11] Wenn man Repräsentate als innere *Werkzeuge* betrachtet, die von Systemen zum Erreichen bestimmter Ziele eingesetzt werden, dann kann man die hier von mir in bezug auf das Phänomen der mentalen Repräsentation eingenommene Position als *Teleofunktionalismus*[12] bezeichnen. Sie deutet auf einen Unterschied zwischen künstlichen und biologischen Repräsentationssystemen hin: Künstliche Systeme – wie wir sie heute kennen – besitzen noch keine Interessen, ihre internen Zustände erfüllen keine Funktion *für* das System – sondern immer nur für die größere Einheit des Mensch-Maschine-Systems. Darum repräsentieren diese Zustände auch nicht im hier intendierten Sinn. Lassen Sie uns nun von diesem kurzen Blick auf die logische Struktur der Repräsentationsbeziehung übergehen zu der Frage, warum sie in den sie realisierenden Systemen die Instantiierung psychologischer Eigenschaften bewirkt.

---

[10] Auf die Bedeutung des Unterschiedes zwischen sequentieller und parallel-distribuierter Informationsverarbeitung für unser Problem komme ich in Abschnitt 2.2 zurück.

[11] Vgl. dazu Eimer 1990, Kapitel 2.

[12] Der Teleofunktionalismus ist der einflußreichste zeitgenössische Versuch, auf eine Reihe von Problemen des klassischen Maschinenfunktionalismus (vgl. Putnam 1975; Block 1978; Block/Fodor 1972) als einer Strategie zur Vereinbarung funktionaler und intentionalistischer Handlungserklärungen (vgl. hierzu Beckermann 1977, 1979) zu antworten. Besonders William Lycan hat wiederholt (zum Beispiel Lycan 1987: Kapitel 5) darauf hingewiesen, daß die funktionalistische Erklärungsstrategie nicht zu einem neurobiologisch unplausiblen *Zwei-Ebenen-Funktionalismus* verkürzt werden darf, weil es eine Kontinuität von Erklärungsebenen gibt. *„Neither living things nor even computers themselves are split into a purely „structural" level of biological/physiochemical description and any one „abstract" computational level of machine/psychological description. Rather, they are all hierarchically organized at many levels, each level „abstract" with respect to those beneath it but „structural" or concrete as it realizes those levels above it. The „functional"/„structural" or „software"/„hardware" distinction is entirely relative to one's chosen level of organization."* (Lycan 1990: 60) Diese Einsicht ist besonders in Zusammenhang mit der Konnektionismus-Debatte und der Modellierung neuronaler Netze von Bedeutung. Der Teleofunktionalismus ist außerdem der Versuch, den zu liberalen Begriff der „Realisierung" des frühen Maschinenfunktionalismus durch Einführung teleonomischer Zusatzkriterien relativ zu einem gegebenen System enger zu fassen und biologisch realistischer zu machen. Vgl. auch Dennett 1969, 1975 und Putnams Dementi in Putnam 1991; weitere Literaturverweise finden sich bei Lycan 1990: 59ff.

Was mentale Repräsentation von Fällen externer Repräsentation oder den eingangs erwähnten Beispielen für *nicht-mentale* interne Repräsentation unterscheidet, ist ein wesentliches Charakteristikum der Resultate des Abbildungsvorgangs: Die durch diesen Vorgang erzeugten internen Datenstrukturen sind „mental" genau deshalb, weil sie die Disposition besitzen bewußt zu werden (vgl. Abschnitt 2.1.4). Mentale Repräsentate sind mental, weil sie bewußt werden können. „Mentalität" möchte ich in diesem Sinne als dispositionale Eigenschaft einer bestimmten Klasse biologisch generierter Datenstrukturen verstehen. Eine Theorie des Mentalen ist also eine Theorie über diejenigen internen Zustände informationsverarbeitender Systeme, die bewußt werden können. Aus diesem Grund darf sich eine solche Theorie nicht auf die Erklärung der kausalen Wirksamkeit und der Genese des intentionalen Gehalts mentaler Zustände beschränken. Die Lösung des Leib-Seele-Problems und die Beantwortung der Frage nach der Intentionalität des Mentalen sind natürlich wichtige Bausteine einer modernen Theorie des Geistes. Doch gerade eine solche moderne, naturalistische Theorie wird in ihrem Kern eine Theorie der Subjektivität sein. Sie darf das psychologische Subjekt nicht eliminieren oder einplanieren, sondern muß uns erklären, wie gewisse interne Datenstrukturen bewußt, subjektzentriert und Träger phänomenaler Qualitäten sein können.

Von primärem Interesse für die philosophische Psychologie sind somit die Resultate gewisser interner Repräsentationsprozesse. Die Prozesse selbst, das heißt die spezifischen Mechanismen, durch die solche Repräsentate erzeugt werden, sind Gegenstand empirischer Einzelwissenschaften. Sind die interessierenden Systeme menschliche bzw. biologische Organismen, so werden es überwiegend die Neurowissenschaften sein, an die wir uns mit Fragen bezüglich des Mechanismus der Repräsentation zu wenden haben. Das mit den Bemühungen der empirischen Wissenschaften verknüpfte philosophische Projekt besteht hingegen zum Großteil in der kommentierenden Ausformulierung einer mit deren Ergebnissen kompatiblen psychologischen Anthropologie.

Das menschliche Nervensystem erzeugt eine Vielzahl von Repräsentaten, die Inhalte subjektiven Bewußtseins niemals zu verändern oder zu beeinflussen scheinen. Das können etwa solche Repräsentate sein, die ein Bestandteil von Regelungs- und Steuerfunktionen für den Stoffwechsel oder das Immunsystem sind. Offensichtlich gibt es jedoch – zum Beispiel beim Phänomen des Somnambulismus oder in hypnotischen Zuständen – auch die Möglichkeit, daß komplexe Repräsentate der Umwelt aktiviert werden, *ohne* daß gleichzeitig phänomenales Bewußtsein erzeugt wird. Außerdem ahnen wir seit Freud, daß unbewußte mentale Zustände unser Verhalten und den Gehalt bewußter mentaler Zustände anscheinend stark determinieren können. Auf denjenigen Aspekt der Subjektivitätsfrage, der mit der „Bewußtheit" mentaler Zustände zusammenhängt, komme ich später zu sprechen. Vorerst kommt es auf ein Zuschreibungskriterium für das Prädikat „mental" an, das unsere *common-sense*-Auffassung wiedergibt, die besagt, daß mentale Zustände wie Gefühle, Meinungen und Wünsche

potentielle Bewußtseinsinhalte sind, also episodische innere Zustände von Personen, die Erlebnischarakter haben oder auch nicht. Was macht eine Teilmenge von durch unsere Gehirne erzeugten Repräsentaten zu mentalen Zuständen, die wiederum die Disposition besitzen, zu subjektiven Erlebnissen zu werden?

Das Kriterium, das ich hier anbieten möchte, lautet: *Introspizierbarkeit*. Mental sind all jene Zustände, die prinzipiell introspizierbar sind, d. h. prinzipiell zu Gegenständen eines zumeist willkürlich eingeleiteten und zielgerichteten Prozesses *innerer Aufmerksamkeit* werden können. Innere Aufmerksamkeit darf aber nicht als Aktivität eines Homunkulus verstanden werden, der den Scheinwerferstrahl seiner ihm immer schon gegebenen Bewußtheit auf verschiedene innere Objekte richtet und sie so zu phänomenalen Individuen[13] macht. Vielmehr ist Introspektion ein in manchen informationsverarbeitenden Systemen ablaufender Vorgang, und zwar eine besondere Variante genau desselben Vorgangs, der das Thema dieses Kapitels ist: Introspektion ist die interne[14] Repräsentation mentaler Repräsentate.

Offenbar ist die interessante Klasse von Repräsentaten genau diejenige, die durch Metakognition selbst wieder zum Gehalt höherstufiger Repräsentate werden kann. Die genaue Auszeichnung dieser Klasse ist eine empirische Frage, bezüglich derer man berechtigte Hoffnungen auf Erkenntnisfortschritte in der näheren Zukunft haben darf. Denn die Fakten, die manche internen Repräsentate zu potentiellen Repräsentanda von Metakognition und somit zu introspizierbaren Zuständen machen, werden durch die Neuroinformatik der Zukunft immer besser beschrieben werden können. „Introspizierbarkeit" ist eine charakteristische Eigenschaft mancher Produkte interner Informationsverarbeitung. Und nicht selten sind wir aus pragmatischen Gründen daran interessiert, internen Zuständen diese Eigenschaft zu verleihen bzw. sie zu aktualisieren: Viele Formen von Psychotherapie versuchen, pathologische mentale Strukturen mit verschiedenen Methoden zu introspizierbaren zu machen, weil sie unter der – sehr starken und häufig theoretisch-argumentativ völlig unabgesicherten – Annahme arbeiten, daß sich diese pathologischen Strukturen allein dadurch, daß sie die Eigenschaft der Introspizierbarkeit gewinnen, auflösen, transformieren oder durch direkte „Verursachung von oben" in ihrer unerwünschten Auswirkung auf das Verhalten und Erleben des Patienten beeinflussen lassen. Außerdem wenden wir alle verschiedene Varianten von Introspektion auch in nicht-therapeutischen Situationen an: Wenn wir un-

[13] Vgl. Lycan 1987, Kapitel 8.

[14] Sie ist nur ein *interner* Repräsentationsprozeß (aber kein *mentaler* Repräsentationsprozeß), weil sie selbst nicht das Potential besitzt, durch Metarepräsentation ihrerseits zum Bewußtseinsinhalt zu werden. Dieser Punkt wird außerhalb des Informationsverarbeitungsansatzes von David Rosenthal in seiner Theorie der *Higher-Order-Thoughts* (vgl. zum Beispiel Rosenthal 1986) und innerhalb einer *Computational Theory of Mind* etwa durch Ray Jackendoff (vgl. Jackendoff 1987) entwickelt.

sere sexuelle Erregung genießen möchten, wenn wir uns konzentrieren, wenn wir uns an etwas zu erinnern versuchen, wenn wir versuchen, herauszufinden, was wir eigentlich *wirklich* wollen oder auch einfach, wenn wir gefragt werden, wie es uns geht. Es gibt zudem passive, nicht ziel- sondern prozeßorientierte Typen von Introspektion, wie den Tagtraum oder verschiedene Formen von Meditation. All diesen Bewußtseinszuständen ist gemeinsam, daß in ihnen mentale Zustände zu Objekten innerer Aufmerksamkeit gemacht werden und ihre Introspizierbarkeit ausgenutzt wird, um sie vorübergehend in den Brennpunkt subjektiven Erlebens zu rücken.

Introspizierbarkeit ist eine notwendige *phänomenologische* Bedingung, wenn interne Systemzustände zu mentalen Zuständen werden und als solche erklärt werden sollen. Phänomenologie allein ist aber nicht genug für eine moderne Theorie des Geistes. Die phänomenologische „Introspizierbarkeit unter Standardbedingungen" liefert uns keinen befriedigenden Begriff der Mentalität, weil sie die hinreichenden Bedingungen seiner Anwendung nicht fixieren kann. Schließlich kennen wir Bewußtseinsinhalte – von unseren Gehirnen erzeugte mentale Modelle von Objekten (d. h. als Objekte codierte Datenstrukturen) in unserer Umwelt –, die wir unter Standardbedingungen niemals als introspizierbar erleben. Die Fortschritte der Neurowissenschaften haben mehr als deutlich gemacht, wie stark auch solche Inhalte in ihrem Auftreten und phänomenalen Gehalt durch die jeweiligen Eigenschaften des Gehirns determiniert sind. Die verschiedenen Typen von Halluzinationen, Agnosien und Neglekten[15] zeigen (ganz unabhängig von jeder theoretischen Position zum Leib-Seele-Problem) wie eng die Korrelation zwischen neuronalen und psychischen Zuständen ist und wie stark die Determination „von unten". Es gibt zum Beispiel Wahrnehmungserlebnisse von externen Objekten, die wir ausgehend von unserem präreflexiven Erleben niemals als „mental" oder „introspizierbar" bezeichnen würden. Die wissenschaftliche Forschung zeigt uns nun aber, daß auch diese Zustände unter abweichenden Bedingungen durchaus als mentale, innere oder „introspizierbare" Zustände *erlebt* werden können.[16] Dies legt einen wichtigen Schluß nahe, der beim gegenwärtigen Stand des empirischen Wissens und aller gebotenen Vorsicht bereits als gerechtfertigt erscheint: Der Prozeß der mentalen Repräsentation generiert in vielen Fällen (zum Beispiel im humanen Wachbewußtsein) sowohl Repräsentate, die aus der Perspektive der ersten Person als mental erlebt werden und potentielle Objekte von Introspektion und innerer Aufmerksamkeit sind, als auch Repräsentate, die als nicht-mental bzw. *externe* Zustände erlebt werden. Mentale Repräsentation – als aus der Dritte-Person-Perspektive des Kognitionswissenschaftlers analysierter Vorgang – führt also in bezug auf die

[15] In Abschnitt 2.3.2 werde ich einige solcher Fälle und ihre Bedeutung für eine naturalistische Theorie mentaler Repräsentation untersuchen.

[16] Das kann zum Beispiel bei Schizophrenien, Manien oder religiösen Erfahrungen der Fall sein.

phänomenale Ebene nicht ausschließlich zu *subjektiven* mentalen Zuständen.[17] Ein Hauptinteresse dieser Arbeit liegt darin, zu verstehen, was es heißt, daß durch Informationsverarbeitung im zentralen Nervensystem manche internen Zustände eben *als* interne bzw. mentale repräsentiert werden, andere dagegen nicht.[18]

Ein weiteres zentrales Problem im Zusammenhang mit dem psychischen Phänomen der mentalen Repräsentation wird durch die Tatsache erzeugt, daß mentale Repräsentate – wie alle physischen Datenstrukturen – in verschiedenen *Formaten* vorliegen. Ich glaube, daß der Begriff des „Formats" für eine Erklärung phänomenaler Zustände unter dem Informationsverarbeitungsansatz eine bisher weitgehend übersehene Bedeutung besitzt.[19] So kann Information zum Beispiel in propositionalem Format vorliegen oder in imaginalem, in Form von Aussagen (bzw. internen Satzanaloga) oder in

[17] Nebenbei bemerkt drückt diese Einsicht eine der vielen Möglichkeiten aus, auf die in einer modernen „informationalistischen" Theorie des Geistes die wesentlichen Einsichten klassischer idealistischer und materialistischer Bewußtseinsphilosophien bewahrt werden können: In einer gewissen Hinsicht ist alles (qua so *Repräsentiertes*) „im Bewußtsein", auch „das Objektive" und die „Widerständigkeit der Welt", aber gleichzeitig sind die zugrundeliegenden Funktionen der Informationsverarbeitung ausschließlich durch physikalische Zustände realisiert.

[18] Die Illusion der Objekthaftigkeit von Gegenständen im subjektiven Bewußtsein, könnte unter dem Informationsverarbeitungsansatz dadurch erklärt werden, daß das Gehirn bei manchen Datensätzen nach der ersten Metarepräsentationsstufe aufhört, den zugrundeliegenden Prozeß zu iterieren. Dahinter steht das Problem, daß von ihrer logischen Struktur her unendliche Prozesse wie Metarepräsentation oder Selbstmodellierung (vgl. hierzu Kapitel 3) durch endliche physikalische Systeme prinzipiell nicht realisiert werden können. Wie ich in meinem **„Pflichtenheft für mentale Repräsentate"** zeigen werde, dürfen Repräsentate das System (wenn es biologisch überlebensfähig sein soll) bei ihrer Konstruktion nicht in infinite Regresse, Endlosschleifen etc. führen. Eine Vermutung wäre, daß das Gehirn eine funktionale Architektur herausgebildet hat, die solche iterativen, aber notwendigen Prozesse wie Metarepräsentation und Selbstmodellierung durch *Objektbildung* abfängt. Analogien zu solchen Phänomenen finden wir in der Logik (vgl. Blau 1986) oder in der Unterscheidung zwischen Objekt- und Metasprache.

[19] Wenn phänomenales Bewußtsein als die Erzeugung interner Datenstrukturen interpretiert werden kann, dann können diese Datenstrukturen unter drei Hinsichten analysiert werden: ihrem *Format*, ihrem *Gehalt* und der Art und Weise, in der sie vom System *organisiert* werden (Vgl. Kosslyn 1981: 47). Will man den intentionalen Gehalt mentaler Zustände unter dem Informationsverarbeitungsansatz untersuchen, dann wird man sich dem zweiten Charakteristikum zuwenden, will man ihre Subjektzentriertheit besser verstehen, dann wird man das zuletzt genannte Merkmal, also die strukturelle Organisation des inneren Repräsentationsraums näher betrachten. Das Format dagegen hängt am direktesten mit der physikalischen Realisierung, dem *Medium* des Repräsentationsprozesses zusammen und schränkt die Anzahl der möglichen Organisationsformen ein – es ist die im Zusammenhang mit Qualia interessanteste informatische Kategorie. *„The format is determined by (a) the nature of the „marks" used in the representation (such as ink, magnetic fluxes, or sound waves) and (b) the way these marks are interpreted (the mark A could be taken as a token of a letter of the alphabet or a picture of a particular pattern). The format specifies wether a representation is composed of primitive elements and relations and, if so, specifies their nature. . . . The format of representation constrains the possible organizations but does not determine them."* (Kosslyn 1981: 47)

Form von Vorstellungsbildern.[20] Das imaginale Repräsentationsformat spielt auch eine wichtige Rolle, wenn wir Thomas Nagels *objective self* besser verstehen wollen. Denn was Nagel von seiner Leserschaft verlangt, ist in Wirklichkeit die Erzeugung gewisser *imaginaler Repräsentate* in Verknüpfung mit einer bestimmten philosophischen Interpretation der dann auftretenden subjektiven Erlebnisse. Man darf bei der Analyse des menschlichen Geistes als eines natürlichen Repräsentationssystems auf biologischer Basis nicht übersehen, daß menschliche Gehirne mit einer Vielzahl von Repräsentatformaten operieren: auditorischen, kinästhetischen, propriozeptiven, gustatorisch/olfaktorischen oder piktorialen. Wir hören *Klänge*, spüren unsere *Körperbewegungen* oder riechen den *Duft* von Gebäck, der vielleicht seinerseits wiederum *innere Bilder* aus unserer Kindheit aufsteigen läßt. Die meisten dieser introspektiv differenzierbaren Formate lassen sich bestimmten Modulen, d. h. vom restlichen Informationsfluß weitgehend abgeschotteten Subsystemen des Gehirns zuordnen: Das Auftreten phänomenaler Bilder ist eng korreliert mit Aktivitäten des visuellen Cortex und Denken im engeren Sinne „internen Sprechens" geht einher mit komplexen Erregungsmustern in den Sprachzentren. Auf die Bedeutung des Formats, in dem eine mentale Datenstruktur intern vorliegt, für ihren qualitativen Gehalt werde ich später zurückkommen.[21]

Interessant ist zudem, daß unser Gehirn nicht nur eine Vielzahl verschiedener Repräsentatformate zu erzeugen vermag, sondern daß es diese auch

[20] Um die Bedeutung von Vorstellungsbildern für die Philosophie des Geistes sind bereits Fachdebatten entbrannt. Pylyshyn hat (1973b, 1979b) kritisiert, daß das Sprechen von inneren „Bildern" nur eine nach dem Modell der äußeren Sinneswahrnehmung gebildete Metaphorik ist, die in innere Regresse und zum Auftreten von *homunculi* führt. Er hat außerdem für einen dritten Code argumentiert, der zwischen satz- und bildartigen Repräsentationsprozessen vermittelt (Eine Analyse verschiedener Typen von Bildhaftigkeit bezüglicher mentaler Repräsentate findet sich bei Rehkämper 1990). Stephen Kosslyn hat in Auseinandersetzung mit dieser Kritk eine empirisch fundierte Theorie mentaler Bilder entwickelt (Kosslyn 1980, 1981, 1983; Kosslyn/Shwartz 1977), die davon ausgeht, daß visuelle Bilder in einem speziellen räumlichen Medium dargestellt werden, das den höchsten Auflösungsgrad in seinem Zentrum besitzt, eine spezielle „Körnung" aufweist und in dem Bilder bald nach ihrer Erzeugung wieder verblassen. Im Langzeitgedächntis finden sich sowohl Bilddateien in analogem Format und propositionale Dateien, die Informationen über Teile der Objekte und Relationen zwischen den Teilen enthalten. Diese beiden Dateitypen werden bei einer Vielzahl von mentalen Operationen miteinander verknüpft (eine knappe Darstellung der Debatte und der neuropsychologischen Belege findet sich in Eysenck/Keane 1990: Kapitel 7). In einer Reihe von Erwiderungen (Pylyshyn 1981, 1984) hat Zenon Pylyshyn der These Plausibilität zu verleihen versucht, daß mentale Bilder *kognitiv penetrabel* (also durch höhere, propositionale Operationen zu beeinflussen und deshalb prinzipiell auf sie reduzierbar) sind und deshalb kein Element der funktionalen Architektur des Systems sein können. Der Begriff der „kognitiven Undurchdringlichkeit" ist im weiteren Verlauf der Debatte vielfach kritisiert worden und auch die simple Unterscheidung zwischen propositionalem und analog-räumlichem Format. Vgl. hierzu Boden 1988, Johnson-Laird 1983. Eine gute, philosophisch orientierte Textsammlung ist von Block 1981 zusammengestellt worden, vgl. auch Block 1983a und Sterelny 1986.

[21] Vgl.Abschnitt 2.1.3.

wieder miteinander verknüpfen und ineinander einbetten kann.[22] Wir sind in der Lage, zu *sehen*, was wir *hören*[23] – d. h. zwei interne Repräsentate, die in zwei unterschiedlichen Gehirnmodulen generiert wurden und in verschiedenen Formaten präsentiert werden, bezüglich ihres Gehalts zu *identifizieren*. Indem man dieser Tatsache Rechnung trägt, stößt man auf ein wichtiges Merkmal der Repräsentate, die für phänomenales Bewußtsein maßgeblich zu sein scheinen: *Multimodalität*. Phänomenale Bewußtseinsinhalte sind überwiegend formatübergreifend und multimodal. Diese Tatsache führt direkt ins Herz einer jeden Theorie, die subjektives Bewußtsein als natürliche Eigenschaft einer bestimmten Klasse von Repräsentationssystemen analysieren will. Auch dieser Punkt wird uns noch beschäftigen.[24]

Ein weiterer, für eine philosophische Theorie des Mentalen zentraler Faktor ist die *Biologizität* mentaler Repräsentate: In unserem eigenen Fall entstanden sie als evolutionäre Strategien, als neue, von gewissen Organismen entwickelte abstrakte Organe. Wir kennen Bewußtsein, Subjektivität und qualitativen Gehalt bisher nur in Zusammenhang mit Systemen, die der biologischen Evolution auf diesem Planeten entstammen. Es ist durchaus denkbar, daß in anderen Teilen des Universums nicht-biologische Repräsentationssysteme entstanden sind, welche ein vergleichbares oder sogar höheres Maß an Komplexität interner Informationsverarbeitung erreicht haben. Was unsere *eigene* Psychologie dagegen angeht, dürfen wir den evolutionären Kontext nicht vergessen, der zur Entwicklung unserer Nervensysteme geführt hat. Die Natur leistet sich nur selten Luxus und deshalb ist es sinnvoll, zu fragen: Welche Vorteile könnte die Entwicklung von Bewußtsein, Subjektivität und qualitativem Gehalt angesichts des erbarmungslosen Selektionsdrucks der biologischen Umwelt auf diesem Planeten gehabt haben?

Will man nun die Suche nach einer Antwort auf die eben gestellte Frage durch eine Analyse der fraglichen biologischen Phänomene im Rahmen einer Theorie mentaler Repräsentation durchführen, dann wird eines der wichtigsten Desiderate in der Entwicklung einer *Typologie* mentaler Repräsentate bestehen. Die drei zentralen Projekte und die ihnen korrespondierenden Fragen sind in diesem Zusammenhang:

– **Das logische Projekt:** Welche Typen von mentalen Repräsentaten sind logisch *möglich*?[25] Was folgt aus der Einführung des Begriffs der mentalen Repräsentation (im Sinne der eben angesprochenen antisymmetrischen Relation) für den des Repräsentats? Mit Blick auf den ontologischen Status von Repräsentaten werden Antworten auf diese Frage den Raum möglicher Leib-Seele-Theorien definieren.

---

[22] Vgl. Abschnitt 2.3.

[23] Vgl. dazu Jackendoff 1987: 300.

[24] Vgl. Abschnitt 2.2.

[25] Eine gute Taxonomie nicht-naturalistischer und gegenwärtiger naturalistischer Theorien intentionalen Gehalts bietet William Lycan in *„Ideas of Representation"*. Vgl. Weissbord 1989.

– **Das psychologische Projekt:** Welche Typen solcher Repräsentate sind *phänomenologisch realistisch?* Welche Eigenschaften müssen sie mindestens besitzen, um in einem Biosystem episodisch eine Psychologie und ein phänomenales Leben zu erzeugen? Gibt es psychologische Eigenschaften, die prinzipiell durch keinen denkbaren Typ mentaler Repräsentation erklärt werden können?
– **Das empirische Projekt:** Welche *Mechanismen* der Informationsverarbeitung sind (in unserem eigenen Fall) neurobiologisch realistisch? Welche Formen interner Repräsentate wären im evolutionstheoretischen Kontext als erfolgreich und damit ausleserelevant anzusehen?
Antworten auf die Fragen des **empirischen** Projekts sind derzeit noch rar – wobei die Geschwindigkeit, mit der sie sich vermehren, gegenwärtig stark ansteigt. Sie müssen und werden jedoch zukünftig die durch Antworten auf das **logische** Projekt definierte Klasse von Repräsentaten extrem einschränken. In den Worten von Patricia Churchland und Terrence Sejnowski:

> To be really useful, a model must be biologically constrained. However, exactly which biological properties are crucial to a model's utility and which can be safely ignored until later are matters that can be decided only by hunches until a mature theory is in place. Such „bottom-up" constraints are crucial, since computational space is immensely vast – too vast for us to be lucky enough to light on the correct theory simply from the engineering bench. Moreover, the brain's solutions to the problems of vision, motor, control, and so forth may be far more powerful, more beautiful, and even more simple than what we engineer into existence. This is the point of Orgel's Second Rule: Nature is more ingenious than we are. And we stand to miss all that power and ingenuity unless we attend to neurobiological plausibility. The point is, *evolution has already done it,* so why not learn how that stupendous machine, our brain, actually works?[26]

Man kann sich dem Ziel einer Typologie mentaler Repräsentate aber vorläufig auch auf dem Umweg über die Beantwortung von Fragen des **psychologischen** Projekts annähern, indem man die *Struktur* der phänomenalen Realität näher analysiert. In diesem Sinn werde ich jetzt versuchen, im Stil eines „Pflichtenhefts für mentale Repräsentate" einige Überlegungen anzustellen. Mein Ziel ist dabei, auf diese Weise zu einem „theoretischen Prototyp", einem idealisierten Begriff *phänomenaler* mentaler Repräsentate zu gelangen, der dann daraufhin überprüft werden kann, ob er empirisch realistisch ist und ob sich theoretische Probleme durch ihn lösen lassen.

Ein Pflichtenheft für mentale Repräsentate

– Die Umwelt (mit „Umwelt" meine ich sowohl das innere als auch das äußere Milieu, in dem ein Organismus operiert) ist dynamisch. Viele ihrer

[26] Vgl. Churchland/Sejnowski 1990(1989): 248; siehe auch Churchland/Sejnowski 1992.

Bestandteile (zum Beispiel andere Lebewesen) verändern ihre Eigenschaften ständig, mit großer Geschwindigkeit und oft unvorhersehbar. Sollen mentale Repräsentate effektive Instrumente im Kampf ums Überleben sein, so müssen sie diese Dynamizität intern abbilden – sie müssen selbst in einem gewissen Ausmaß *dynamische Repräsentate* sein. Das bedeutet: Sie müssen unabhängig von der Umgebung des Systems bzw. dem Strom des Inputs aktivierbar sein.[27]

– Die Umwelt von Biosystemen besteht zu einem nicht unbeträchtlichen Teil aus Kontinua. Der Säuregehalt des Wassers oder der Adrenalinspiegel im Blut[28] sind allgemeine Hintergrundparameter, deren Schwankungen lebenswichtig sein können. Für solche analogen Signale aus der Umwelt bzw. den Sinnesorganen muß es eine Möglichkeit der *Intensitätsdarstellung* geben. Das bedeutet, daß zumindest manche mentalen Repräsentate in der Lage sein sollten, die wechselnde Intensität oder Signalstärke abzubilden, mit der die jeweiligen Repräsentanda die Transduktoren beeinflussen. Ich werde für solche funktionalen Zustände in Abschnitt 2.1.3 den Begriff des *Analog-Indikators* einführen.

– Um höhere kognitive Vorgänge einleiten zu können, müssen auch Abstrakta wie Klassen oder Relationen abgebildet werden. Das bedeutet: Es muß umfassende Repräsentate geben, die Datenstrukturen aus verschiedenen Modulen und Subsystemen des Gehirns integrieren, wie sie etwa den Sinnesorganen nachgeschaltet sind. Soll der phänomenale Gehalt unserer Erlebnisse vollständig auf mentale Repräsentation zurückgeführt werden, dann muß es multimodale Repräsentate geben. Um höhere Abstraktionsleistungen (zum Beispiel Verallgemeinerungen oder Begriffsbildungen) zu erbringen, müssen aus Datenstrukturen verschiedener Herkunft und unterschiedlichen Formats *Prototypen* erzeugt werden.

– Die Notwendigkeit umfassender Repräsentate bedingt die Postulierung rekursiver Funktionen[29], die mentale Repräsentate ineinander *einbetten*. Die empirische Frage lautet hier, ob es Hinweise auf solche Funktionen und die Art ihrer neurobiologischen Realisierung gibt. Aus theoretischer Perspektive muß einsichtig gemacht werden können, wie Teile der phänomenalen Wirklichkeit psychisch individuiert werden und wie sie zu Elementen größerer Teilbereiche bei gleichzeitiger Homogenität des repräsen-

---

[27] Auf die Frage der input-unabhängigen Aktivierbarkeit mentaler Repräsentate gehe ich in Abschnitt 2.1.2 ein. Was uns derzeit noch von künstlichen Systemen unterscheidet, ist die Tatsache, daß wir über eine autonome innere repräsentationale *Dynamik* verfügen.

[28] Der Adrenalinspiegel ist hier als Repräsentandum gedacht. Interessant ist jedoch, daß er auch als *Mechanismus* der internen Repräsentation gesehen werden kann. Es gibt nicht-stationäre Systeme der Infomationsübertragung und Verhaltenssteuerung wie das Immunsystem oder die Hormone, die für den Gehalt mancher mentaler Repräsentate mindestens eine ebenso große – und bis heute weitgehend unverstandene – Rolle spielen wie das Feuern einzelner Neuronen oder die Aktivierung von Zellverbänden.

[29] Johnson-Laird hat auf die Notwendigkeit solcher Funktionen hingewiesen. Vgl. Johnson-Laird 1983, 1989.

tationalen Gesamtzustandes[30] werden können. Diese Eigenschaft der wechselseitigen Einbettbarkeit mentaler Repräsentate besitzt (wie wir noch sehen werden) große Bedeutung für Subjektivität als einer relationalen Eigenschaft.

– Mentale Repräsentate sind *gute* Repräsentate, wenn sie für den Evolutionserfolg wichtige kausale Eigenschaften der Systemumwelt erfolgreich und zuverlässig abbilden. Darum habe ich auch sie als von Biosystemen entwickelte abstrakte Organe bezeichnet: Mentaler Gehalt wird zum *Instrument* eines Systems. Die Umwelt besteht aber aus einem extrem komplexen Netzwerk kausaler Relationen, von denen ein Teil direkt überlebensrelevant ist. Deswegen muß diese relationale Struktur sowohl einzelner Repräsentanda als auch der biologischen Umwelt als Ganzer intern repräsentiert werden. Mentale Repräsentate müssen systematisch mit anderen Repräsentaten verknüpft sein; sie müssen eine *partielle relationale Homomorphie* zu ihren Repräsentanda besitzen, wenn sie biologisch nützlich sein sollen.

– Bei primitiven Biosystemen mag es so sein, daß es nur interne *Präsentation*[31] gibt. Solche Organismen wären in einem ewigen Jetzt gefangen, der Gehalt ihrer mentalen Zustände bezöge sich nur auf aktuell gegebene Reizquellen. Will man dagegen Gedächtnis und höhere kognitive Leistungen wie die interne Repräsentation von Zielzuständen („Zukunftsplanung") als biologische Phänomene erklären, dann muß man für das System eine Möglichkeit annehmen, komplexe mentale Repräsentate unabhängig vom dauernden Strom des Inputs[32] zu erzeugen. Da das System hierzu nicht oder nur begrenzt auf den Informationsfluß aus den Sinnesmodulen zugreifen kann, wird es sich dabei die eben erwähnte relationale Struktur der mentalen Repräsentate zunutze machen. Das heißt: Mentale Repräsentate müssen interne Simulationen komplexer, kontrafaktischer Situationen ermöglichen. Auf der phänomenalen Ebene finden sich reichhaltige Beispiele für mentale Simulationen[33] – man denke an sexuelle Phantasien, das Grübeln

[30] Repräsentationale Gesamtzustände werden durch die Gesamtheit aller zu einem gegebenen Zeitpunkt bewußt erlebten mentalen Modelle gebildet. Sie bestehen aus dem aktuellen Weltmodell und dem aktuellen Selbstmodell und bilden das gegenwärtige *Realitätsmodell* des Systems. Repräsentationale Gesamtzustände sind Zustände von Realitätsmodellen, d. h. von bewußt erlebten Makro-Repräsentaten. Kann man für ein gegebenes System eine Reihe von diskreten, strukturell definierten Makro-Repräsentaten unterscheiden, so kennzeichnet man Elemente der Menge der diesem System möglichen Bewußtseinszustände. Vgl. Abschnitt 2.3.1.

[31] Vgl. Abschnitt 2.1.3.

[32] Daß dies ein Problem darstellt, sieht man schon daran, daß es tagsüber schwieriger ist, zu träumen, als nachts. Realitätsmodelle sind desto plastischer, je weniger sie durch den aktuellen Input und die funktionale Architektur des Systems determiniert werden. Das Beispiel des Traums zeigt, daß Plastizität häufig Instabilität und niedrigen epistemischen Gehalt bedeutet.

[33] Nach der oben angebotenen Definition von mentaler Repräsentation sind mentale Simulationen keine Fälle von Repräsentation, da es für sie keine aktuellen Repräsentanda gibt. Mentale Repräsentate können also sogar zu Instrumenten in nicht-repräsentationalen Situationen werden, zum Beispiel bei der Zukunftsplanung. Vgl. Abschnitt 2.1.2.

über einem Problem[34], Tagträume (etwa die *subjektive* Erzeugung objektiver Selbste) oder scheinbar nutzlose innere Monologe. Vielleicht kann man sich den spezifischen Vorgang interner Simulation mit Hilfe mentaler Repräsentate so vorstellen, daß ein bestimmtes interessantes Repräsentat probeweise in das aktuelle Gesamtmodell der Welt eingebettet wird. Durch Metarepräsentation[35] kann dann festgestellt werden, wie sich durch diese Einbettung einer neuen relationalen Struktur der Gehalt anderer Repräsentate verändert. Damit solche internen Simulationen realistisch und biologisch erfolgreich sind, müssen sie in den meisten Fällen ein kritisches Ausmaß von Strukturgleicheit mit der Zielsituation erzeugen.[36] Dies wiederum hängt davon ab, daß mentale Repräsentate möglichst viele der überlebensrelevanten relationalen Eigenschaften ihrer externen Repräsentanda durch ihre *eigene* relationale Struktur wiedergeben.

– Obwohl mentale Repräsentate wechselseitig ineinander einbettbar sein müssen, dürfen sie das System nicht über iterative Prozesse in Endlosschleifen führen und dadurch paralysieren. Die Natur muß infinite Regresse vermeiden, und darin mag der Grund für die vielleicht interessanteste Eigenschaft mentaler Repräsentate liegen: Sie sind uns fast nie *als solche* introspektiv gegeben, sie sind selbst-referentiell opak. Das bedeutet: Sie werden so zuverlässig und schnell aktiviert, daß wir – um einen Ausdruck von Robert van Gulick zu verwenden[37] – normalerweise *durch sie hindurchschauen*. Obwohl diese Metapher einen Homunkulus ins Leben ruft, ver-

---

[34] Nachdenken könnte die interne *Simulation der externen Manipulation physikalisch diskreter Symbol-tokens* sein. Vgl. Abschnitt 2.2.1, Bechtel/Abrahamsen 1991, Helm 1990, Goschke/Koppelberg 1990: 267. Johnson-Laird liefert empirische Belege für die Rolle mentaler Modelle während des inferentiellen „Nachdenkens" über ein Problem. Vgl. Johnson-Laird 1983, 1989.

[35] Dies könnte eine evolutionär relevante Rolle von Bewußtsein gewesen sein. Vgl. Abschnitt 2.1.4.

[36] Für Meinungen hat Peter Bieri darauf hingewiesen, daß aus ihrer funktionalen Adäquatheit nicht ihre Wahrheit folgt, weil sogar falsche Meinungen *funktional adäquat* sein können. *Mutatis mutandis* gilt dies auch für mentale Simulationen: Illusionen etwa bezüglich der Eigenschaften intern simulierter Zielzustände können *biologisch adäquat* sein. Vgl. Bieri 1987a: 137, 1987b: 61ff.

[37] Vgl. Van Gulick 1988: 178. Van Gulick bezeichnet diese Eigenschaft auch als *„semantische Transparenz"*: Der intentionale Gehalt ist meist so schnell und deutlich gegeben, daß wir meinen, es mit dem Repräsentandum direkt zu tun zu haben. Es scheint so zu sein, daß die phylogenetisch ältesten, zuverlässigsten und schnellsten Repräsentate (etwa aus den Sinnesmodulen) auch den höchsten Grad an semantischer Transparenz aufweisen, d. h. introspektiv am schwersten als *Konstrukte* zu entlarven sind. – Man darf den Gulickschen Begriff der „semantischen Transparenz" nicht mit dem von Andy Clark verwechseln. Clark möchte mit ihm klassische kognitivistische Systeme im Sinne von Fodor und Pylyshyn auszeichnen, die eine interne Syntax besitzen und eine regelgeleitete Transformation von Symbolen durchführen, welche auf formalen Eigenschaften mentaler Repräsentate beruht. Die Einführung des Terms „semantisch transparentes System" (STS) geschieht hier – in Anlehnung an Arbeiten von Smolensky; vgl. Smolensky 1987, 1988 – mit der Zielsetzung der *Abgrenzung* solcher Systeme gegenüber konnektionistischen Systemen, die repräsentationalen Gehalt intern durch „verborgene Einheiten" darstellen. Vgl. Clark 1989: 17–21, 111–120.

deutlicht sie doch, wie sehr effektive Repräsentate introspektiv nicht mehr als solche erkennbar sein könnten.

Fassen wir kurz zusammen: Soll der Gehalt phänomenaler Zustände als Produkt mentaler Repräsentation analysiert werden, so muß die interessante „prototypische" Klasse von Repräsentaten sich durch Dynamizität, Analogizität[38], Multimodalität bzw. Formatintegration, wechselseitige Einbettbarkeit, partielle Identität der relationalen Struktur mit der ihrer Repräsentanda durch geeignete Verknüpfung mit anderen Repäsentaten, durch Simulationsfähigkeit und durch die van Gulicksche „semantische Transparenz" auszeichnen. Daß ein mentales Repräsentat zu dieser hypothetischen Klasse[39] phänomenaler Repräsentate gehört, heißt nicht, daß es sich durch *alle* diese Eigenschaften auszeichnen muß. So ist zum Beispiel die Eigenschaft „der semantischen Transparenz" am stärksten ausgeprägt bei mentalen Repräsentaten aus den Sinnesmodulen. Hier sind die phänomenalen Qualitäten der Konkretheit und Objekthaftigkeit am prägnantesten: Es ist wesentlich schwieriger, das Buch, das Sie gerade in ihren Händen halten, als intern generierten Zustand zu erkennen, als die beim Lesen entstehenden Gedanken und Gefühle. Das liegt nicht nur daran, daß Ihr visueller Cortex dieses Buch – unter den Standardbedingungen des nichtpathologischen Wachzustandes – *als* externes Objekt repräsentiert. Die Ursache für dieses mentale Phänomen liegt auch in der Geschwindigkeit der jeweiligen Hirnfunktionen.[40] Beim Menschen arbeiten die (stammesgeschichtlich älteren) Sinnesmodule wesentlich schneller und effizienter als etwa die Sprachzentren. Aus diesem Grund ist es – was sinnliche Wahrnehmungen angeht – für das Gehirn fast unmöglich, den zur Aktivierung eines mentalen Repräsentats führenden Konstruktionsprozeß *selbst* noch einmal mental zu repräsentieren. Höhere mentale Vorgänge wie zum Beispiel die episodisch auftretenden „Gedankenketten" dagegen zeichnen sich auch subjektiv durch Prozessualität[41] aus: Wenn wir denken, *geschieht* etwas – und zwar *in* uns. Auch der Gehalt unserer Gedanken verändert sich noch,

[38] Zum Begriff der „analogen Repräsentation" in bezug auf die menschliche Wissensdarstellung und die ihm entsprechenden empirischen Befunde vgl. Steiner 1988.

[39] Es ist natürlich klar, daß die genauere Auszeichnung dieser Klasse eine primär empirische Frage ist.

[40] Gehaltskonstanz von Repräsentaten ist das Ergebnis dynamischer Musterwiederholung. Ist die Frequenz der Musterwiederholung zu hoch für das zeitliche Auflösungsvermögen der jeweiligen Metarepräsentationsfunktion, dann entsteht – ähnlich wie beim Fernsehbild – auf der phänomenalen Ebene eine Kontinuitätsillusion.

[41] Die Prozessualität des Mentalen ist eines der größten Probleme für behavioristische Dispositionsanalysen à la Ryle: Dispositionen besitzen keinerlei Prozeßcharakter. Vgl. Ryle 1949. Die Prozessualität mancher mentaler Inhalte geht auch in Beschreibungen durch Programm-Listings oder Turing-Maschinentafeln verloren. Für konnektionistische Systeme dagegen läßt sich ein präzise Beschreibung des Prozesses geben, durch den ein lernendes System sich unter den jeweils gegebenen Einschränkungen in seinen energieärmsten Zustand begibt und dadurch den Input *interpretiert*: Als Trajektorie durch den Zustandsraum (bei der Betrachtung mehrerer Lernzyklen als Trajektorie im *Gewichtungsraum*).

nachdem sie zu Inhalten subjektiven Bewußtseins geworden sind. Gedanken „nehmen Form an“ und „konkretisieren sich“, d. h. sie werden *als Konstrukte* erlebt, weil der Konstruktionsprozeß partiell mental mitrepräsentiert wird. Ist dieser Prozeß – wie in den meisten Fällen sinnlicher Wahrnehmung – zu schnell für das zeitliche Auflösungsvermögen der metarepräsentierenden Funktion, dann erscheinen die jeweiligen Inhalte als „gegeben“ und nicht als erzeugt.[42] Auf die mentale Qualität der „Gegebenheit“ werde ich im übernächsten Abschnitt noch einmal kurz zurückkommen, indem ich einen Sonderfall mentaler Repräsentation betrachte. Zunächst möchte ich jedoch die Aufmerksamkeit meiner Leser auf die Tatsache hinlenken, daß Repräsentation selbst nichts weiter ist als der Sonderfall eines wesentlich umfassenderen psychischen Phänomens. Zu diesem Zweck müssen wir noch einmal kurz zur logischen Struktur der von subjektiven Erlebnissen begleiteten mentalen Operationen zurückkehren.

2.1.2 Mentale Simulation: Die Erzeugung virtueller Erlebniswelten durch Phantasie, Imagination und Planung

Mentale Repräsentate sind von Gehirnen benutzte Werkzeuge. Diese Werkzeuge werden von biologischen Systemen eingesetzt, um möglichst viel überlebensrelevante Information möglichst schnell und effektiv zu verarbeiten. Ich habe den Prozeß, durch den sie erzeugt werden, als eine dreistellige antisymmetrische Relation zwischen ihnen, einem System und externen oder internen Repräsentanda analysiert. Es fällt sofort ins Auge, daß es viele Fälle gibt, in denen diese Analyse offenkundig falsch ist. Denn es ist ein wichtiges Charakteristikum menschlichen phänomenalen Bewußtseins, daß mentale Repräsentate oft auch dann aktiviert und miteinander verknüpft werden, wenn die ihren Gehalt bildenden Zustände der Welt keine aktuellen Zustände sind: Gehirne können *mögliche* phänomenale Welten erzeugen.

Die dem Entstehen möglicher phänomenaler Welten zugrundeliegenden „virtuellen Repräsentationsprozesse“ erzeugen subjektive Erlebnisse, die den aktuellen Zustand der Welt nicht oder nur teilweise widerspiegeln. Beispiele für solche Zustände sind spontane Fantasien, innere Monologe, Tagträume, Halluzinationen oder nächtliche Träume. Sie umfassen aber auch absichtlich eingeleitete kognitive Operationen: die Planung möglicher Handlungen, die Analyse zukünftiger Zielzustände aller Art, das absichtli-

[42] Wenn der „Bewußtheits“-Aspekt des Subjektivitätsproblems sich über den Begriff der „Meta-Repräsentation“ lösen läßt, dann gilt natürlich auch für Metarepräsentation als Ganzes: Das durch sie erzeugte Bewußtsein kann nur als gegeben erlebt werden, weil *ex hypothesi* der verantwortliche Mechanismus „im Dunklen“ bleiben muß. Dieser Punkt wird auch in David Rosenthals Theorie der *Higher Order Thoughts* und bei Ray Jackendoff deutlich: Der *Computational Mind* ist umfassender als der *Phenomenal Mind* und seine Entdeckung erzeugt eine verfeinerte Variante des Leib-Seele-Problems: das *Mind-Mind-Problem.* Vgl. Rosenthal 1986 und Jackendoff 1987.

che „vergegenwärtigen“ vergangener mentaler Zustände usw. Offensichtlich haben wir es hier nicht mit einem Fall von mentaler Repräsentation zu tun, weil die jeweiligen Repräsentanda nicht oder nur teilweise als Elemente der aktuellen Systemumwelt gegeben sind. Der erste Schluß, den man aus dieser Beobachtung ziehen kann, lautet: Die an den fraglichen mentalen Operationen beteiligten Repräsentate werden *ohne Input* aktiviert. Interessanterweise erleben wir die oben angesprochenen Phänomene häufig dann, wenn die Verarbeitungskapazität des Gehirns nicht besonders stark beansprucht wird, weil keine neuen und schwierigen praktischen Problem zu lösen sind (bei Routinetätigkeiten; zum Beispiel, wenn wir wieder einmal mit dem Auto im Stau stehen) oder weil die aus der Umwelt einströmende Information stark abnimmt (in Ruhephasen, beim Einschlafen). Es muß also einen mehr oder weniger unspezifischen internen Aktivierungsmechanismus geben, der die physischen Randbedingungen für solche Zustände schafft.[43] Ich werde solche mentalen Zustände, die kontrafaktische Situationen repräsentieren, ab jetzt als m*entale Simulationen* bezeichnen.

Mentale Simulation: $\mathbf{M}_{\mathbf{Sim}}$ (S,X,Y)
- X simuliert Y *für* S.
- X ist ein mentales Simulat; d. h. es ist das mentale Repräsentat einer kontrafaktischen Situation.
- Y ist eine kontrafaktische Situation.
- das teleologische Zusatzkriterium muß nicht erfüllt sein; die Aktivierung mentaler Simulate kann auch durch zufällige Mikro-Ereignisse auf der neuronalen Ebene ausgelöst werden (Traum, Halluzination).
- $\mathbf{M}_{\mathbf{Sim}}$ ist antisymmetrisch und kein Fall von $\mathbf{M}_{\mathbf{Rep}}$.

Mentale *Präsentate*[44] können im allgemeinen nicht zu Simulaten werden[45], weil sie in ihren physischen Randbedingungen an einen konstanten Inputstrom gebunden sind – sie können nicht *re*-präsentiert werden. Mentale Repräsentate dagegen können auch außerhalb der Standard-Reizsituationen aktiviert und für mentale Operationen eingesetzt werden: Sie verlieren

[43] Ein Kandidat für dieses unspezifische Aktivierungssystem ist der älteste Teil unseres Gehirns: die *Formatio reticularis*, der Kern des Hirnstamms. Sie kann elektrokortikale Rhythmen aktivieren und desynchronisieren.

[44] Vgl. Abschnitt 2.1.3.

[45] Ausnahmen bilden all jene Situationen, in denen das System mit *internen* Signalquellen einer ausreichenden Stärke konfrontiert ist, wie im Traumzustand oder bei Halluzinationen. Ich diskutiere einige solcher Situationen in Abschnitt 2.3.2.

ihren intentionalen Gehalt[46], behalten jedoch einen Großteil ihres *phänomenalen* Gehalts und werden dadurch zu mentalen Simulaten. Wenn das richtig ist, müßte imaginalen Repräsentaten – wie zum Beispiel Vorstellungsbildern – der qualitative „Signalaspekt" fehlen, der Präsentate auszeichnet. Und in der Tat sind für die meisten von uns absichtlich vorgestellte Schmerzen nicht wirklich *schmerzhaft* und imaginierte Erdbeeren nicht wirklich *rot*[47] (außer wir sind von Geburt an Eidetiker oder haben unser Gehirn durch Visualisierungsübungen trainiert). Zudem werden bei absichtlich eingeleiteten mentalen Simulationen die phänomenalen Qualitäten der „Gegebenheit" und „Instantanität" in wesentlich schwächerem Ausmaß erzeugt.

Organismen, die Simulate nicht *als solche* erkennen und für Repräsentate bzw. Präsentate halten, träumen oder halluzinieren. Und tatsächlich entstehen solche mentalen Zustände sehr oft dadurch, daß durch die Enthemmung gewisser Hirnbereiche starke interne Signalquellen ins Leben gerufen werden. In solchen Situationen ist es dem menschlichen Gehirn allem Anschein nach nicht möglich, diese Reize *als interne* zu repräsentieren. Deswegen haben wir in Träumen, bei Psychosen oder unter dem Einfluß psychoaktiver Substanzen auch manchmal *wirklich* Angst.[48] Andererseits gibt es auch pathologische Zustände, in denen Repräsentate als *Simulate* erscheinen und die Unterscheidung zwischen phänomenalem Innen und phänomenalem Außen auf die umgekehrte Weise verschwindet.

Es erscheint in diesem Zusammenhang wichtig, die genetische und die logische Dimension des Phänomens „mentale Simulation" deutlich zu trennen. Die Entwicklungsgeschichte mentaler Zustände von rudimentären Urformen zu immer komplexeren Makro-Repräsentaten, deren Aktivierung die Instantiierung immer neuer und höherer psychologischer Eigenschaften nach sich zieht, war eine *biologische* Geschichte. Unter dem Selektionsdruck der biologischen Umwelt entstanden immer neue und immer erfolgreichere Formen mentalen Gehalts.[49] Vielleicht kann man die Genese

[46] Das soll heißen: Sie stellen für das System nicht mehr die *wirkliche* Welt dar. Wenn unsere Ontologie Abstrakta wie zum Beispiel mögliche Welten umfaßt, dann können wir hier auch weiterhin von einer echten Repräsentationsbeziehung sprechen – falls wir angeben können, was in diesem Fall die „Standard-Kausalketten" sind.

[47] Man kann sagen: Imaginierte Schmerzen und Roterlebnisse sind *unterdeterminierte* oder nicht vollständig individuierte mentale Entitäten.

[48] Eine Ausnahme bilden Zustände, in denen das System Simulate *als solche* klassifiziert, der Zustand aber trotzdem weiter andauert. Beispiel für solche repräsentationalen Situationen in denen *Zustandsklarheit* herrscht, obwohl das System von Artefakten überschwemmt wird sind Pseudo-Halluzinationen (vgl. Abschnitt 2.3.2) und Klarträume (vgl. Abschnitt 3.2.2).

[49] Viele Autoren haben auf die biologische Funktionalität von mentalem Gehalt hingewiesen. Colin McGinn betont, daß das, was er in Anlehnung an Ruth Millikan die *„relationale Eigenfunktion"* repräsentationaler mentaler Zustände nennt, mit ihrem extrinsisch individuierten Gehalt koinzidiert (zum Beispiel 1989: 147), d. h. daß die Relationalität von mentalem Gehalt die Relationalität des begleitenden biologischen Zustands reflektiert. Diese Sichtweisen sind der von mir in diesem Kapitel entwickelten Perspektive auf mentale Modelle als abstrakten Organen eng verwandt. Vgl. McGinn 1989; außerdem Churchland, P.S. 1986, Dretske 1986, Fodor 1984, Millikan 1984, Papineau 1987, Stich 1992.

komplexer mentaler Repräsentate sinnvollerweise als die biologische Geschichte gewisser innerer Zustände beschreiben, die im Laufe der Zeit ein immer größeres Maß an interner Relationalität und Autonomie im Sinne von Inputunabhängigkeit erreicht haben.

Die erste Form von komplexer Reizverarbeitung und intelligenter Interaktion mit der Umwelt mag der *Reflexbogen* gewesen sein: Ein fest verdrahteter Weg vom Reiz zu einer starren motorischen Reaktion ohne Erzeugung eines spezifischen inneren Zustandes. Die nächste Stufe war das *mentale Präsentat*. Es zeichnet sich durch mehr oder weniger ausgeprägte Output-Entkopplung aus. Das heißt: Mentale Präsentate sind spezifische innere Zustände, die die *aktuelle Präsenz* eines bestimmten Sachverhalts bezüglich der Welt oder des Systems anzeigen. Sie sind neue mentale Instrumente, die es einem Organismus erstmals ermöglichen, Information intern zu präsentieren, ohne sofort in einer festgelegten Weise auf sie reagieren zu müssen. Ihr Nachteil besteht in ihrer Inputabhängigkeit: Da sie nur durch kontinuierlichen Input aufrechterhalten werden können, bilden sie auch nur die aktuelle Präsenz einer Reizquelle ab. Um ein weiteres Mal zu dem klassischen Beispiel zurückzukehren: Ein Schmerzerlebnis *präsentiert* dem Erlebnissubjekt einen Gewebeschaden oder eine anderweitige körperliche Verletzung. Bis zu einem gewissen Intensitätsgrad dessen, was ich als den „Signalaspekt" bezeichnet habe, muß die betreffende Person überhaupt nicht mit Verhalten reagieren. Aber selbst wenn sie durch die Stärke des „puren Präsentationsaspekts" zu einer Reaktion gezwungen ist, kann sie aus einer mehr oder weniger großen Palette möglicher Verhaltensweisen auswählen. Der Nachteil von Schmerzen ist, daß wir uns ihr volles subjektives Erlebnisprofil nur schwer *vergegenwärtigen* können, nachdem sie abgeklungen sind.

Der wesentliche Übergang zur Erzeugung einer genuinen inneren Wirklichkeit wird dann in der Input-Entkoppelung mancher Zustände bestanden haben. Nun konnten Relationen (zum Beispiel kausale Relationen) zwischen Repräsentanda auch dann intern abgebildet werden, wenn die Repräsentanda nicht oder nur teilweise in Form der typischen Reizquellen gegeben waren. Dadurch entstand die Möglichkeit, abstrakte Information zu verarbeiten. Interne Simulate haben natürlich auch andere subjektive Eigenschaften als Präsentate, weil sie eine andere kausale Geschichte durchlaufen haben. Sie können in umfassendere Repräsentate eingebettet werden und auch dann aktiviert werden, wenn ihr Repräsentandum nicht durch den Inputstrom gegeben ist, sondern über die relationale Struktur anderer Repräsentate. Repräsentate können sich untereinander aktivieren, weil sie über ihre physischen Entstehungsbedingungen miteinander vernetzt sind.[50]

[50] Innerhalb konnektionistischer Systeme (zum Beispiel in neuronalen Netzen) kann diese assoziative Verknüpfung interner Repräsentate über ihre Ähnlichkeit bzw. über ihre Lage in der durch das System gebildeten „Energielandschaft" erklärt werden. Die *repräsentationale* Ähnlichkeit von Aktivierungsvektoren findet also ihren *physikalischen* Ausdruck in der Wahrscheinlichkeit, mit der zwei stabile Aktivierungszustände des Systems gleichzeitig auftreten können.

So ist es möglich, daß höherstufige mentale Strukturen entstehen, deren repräsentationaler Gehalt nicht oder nur teilweise aus zum Zeitpunkt ihres internen Auftretens ebenfalls gegebenen Fakten besteht. Diese höherstufigen mentalen Strukturen lassen sich am besten durch ihre *Funktion* verstehen: Sie ermöglichen dem Organismus die interne Simulation komplexer kontrafaktischer Ereignisfolgen. Dadurch werden höhere kognitive Leistungen wie Gedächtnis und strategische Zukunftsplanung ermöglicht. Die neuen Werkzeuge, mit denen diese Leistungen erbracht werden, sind mentale Simulate – innere Zustände, die sich die relationale Vernetzung mentaler Repräsentate untereinander zunutze machen, um unabhängig von äußerem Input umfassendere innere Strukturen zu aktivieren. Diese Simulate können dann zum Beispiel mit Sollzuständen verglichen werden. *Präsentate* ermöglichten durch Output-Entkoppelung die Entwicklung eines größeren behavioralen Repertoires zu einer gegebenen Reizsituation. *Re*präsentate gestatteten durch weitgehende Input-Entkoppelung die Entwicklung eines größeren *inneren* behavioralen Repertoires, wenn sie durch interne Ursachen – also: *als Simulate* – aktiviert wurden. Man kann sagen: Mentale Simulation ist eine neue Form von Verhalten, von *innerem Handeln*.[51] Die Erzeugung komplexer mentaler Simulate, die in einem gewissen Ausmaß unabhängig ist vom Strom des aktuellen Inputs und nicht zwangsläufig zu motorischem Makro-Verhalten[52] führt, ist die Vorbedingung für diese neue Form des Verhaltens.

Dies könnte in groben Zügen die biologische Geschichte der Entwicklung komplexer interner Zustände gewesen sein, die schließlich Repräsentationalität und Funktionalität in sich vereinten.[53] Vielleicht reicht der Begriff des „Präsentats", den ich im folgenden Abschnitt kurz erläutern werde, bereits aus, um eine erfolgversprechende Perspektive auf das Qualia-Problem einzunehmen, jedoch sind die theoretisch vertracktesten psychischen Phänomene wie Bewußtsein und relationale Subjektivität durch die oben skizzierte biologische Genese noch nicht zu verstehen.

---

[51] Höhere kognitive Leistungen wie die Bildung von Theorien oder die Planung zielgerichteten Handelns werden deshalb überhaupt nur mit solchen inneren Werkzeugen möglich, die nicht mit Eigenschaften der Umwelt kovariieren: Der Gehalt und der Erfolg kognitiver Modelle kann nicht durch eine Kovarianztheorie erklärt werden. *„But, in order to model possible worlds, we must have cognitive models able to break away from covariance with the actual world. .. If we are going to treat all cases of non-covarying representation as cases of ‚mis'representation, then it seems that misrepresentation is by no means sub-optimal, but is in fact a necessary and integral part of cognition."* Kukla 1992: 222. Vgl. Kukla 1992, Dretske 1981, Cummins 1989.

[52] Der Somnambulismus ist ein Fall, bei dem die funktionale Entkoppelung der Simulation von der Motorik nicht gelingt.

[53] Wenn diese Vermutung in die richtige Richtung weist, deutet sie auch eine Lösung des Leib-Seele-Problems an. Die Frage nach der *Nomizität von mentalem Gehalt* können wir dann beantworten, wenn wir eine theoretische Entität dingfest gemacht haben, der sowohl verhaltensauslösende funktionale und physikalische Zustände des Systems, als auch repräsentationale Träger von Gehalt entsprechen. Aktivierungsvektoren in konnektionistischen Systemen sind gute vorläufige Kandidaten für die Besetzung dieser Schlüsselposition in unserer Theorie des Geistes – Subjektivität werden wir mit ihrer Hilfe allein nicht verstehen können.

Wenden wir uns vorerst noch einmal kurz der logischen Dimension des Phänomens der mentalen Simulation zu. Was bedeutet es, daß $\mathbf{M}_{\mathbf{Sim}}$ kein Fall von $\mathbf{M}_{\mathbf{Rep}}$ ist? Was ist damit gesagt, daß durch den Prozeß der mentalen Simulation in einem System kontrafaktische Sachverhalte abgebildet werden? Man kann mentale Repräsentation als einen *Sonderfall* von mentaler Simulation rekonstruieren, nämlich als genau den Fall von mentaler Simulation, bei dem **erstens** das Simulandum zum Zeitpunkt t der Aktivierung des internen Simulats als *Repräsentandum*, d. h. als Bestandteil der wirklichen Welt, gegeben ist; und **zweitens** das Simulandum die Aktivierung des Simulats durch die Standard-Kausalketten auslöst.

Mentale Simulation: $\mathbf{M}_{\mathbf{Sim}}$ (S,X,Y)

- Es gibt eine *mögliche* Welt W, so daß $\mathbf{M}_{\mathbf{Sim}}$ (S,W,X,Y) und Y ein Element von W ist.

Mentale Repräsentation: $\mathbf{M}_{\mathbf{Rep}}$ (S,X,Y) = $\mathbf{M}_{\mathbf{Sim}}$ (S,W_0,X,Y),

- so daß Y zum Zeitpunkt t ein Element der *wirklichen* Welt W_0 ist, also:
- Y ist ein erfüllter Sachverhalt in W_0;
- Y löst zum Zeitpunkt t X in S über die Standard-Kausalketten aus.

Nach diesem Schema ist jede Repräsentation *auch* eine Simulation, da es – mit der wirklichen Welt – mindestens *eine* mögliche Welt gibt, in der das Repräsentandum einen Sachverhalt darstellt. Der Gehalt mentaler Simulate besteht somit aus Sachverhalten in möglichen Welten.[54] Von der logischen Struktur her gesehen ist also Simulation das umfassendere Phänomen und Repräsentation nur ein eingeschränkter Sonderfall. Dagegen ist aus genetischer Perspektive die Repräsentation das *frühere* Phänomen, weil Organismen überhaupt erst durch Wahrnehmung der Umwelt die funktionale Architektur ihrer Nervensysteme geprägt haben, die sie sich dann später auch zur nicht-repräsentationalen Aktivierung mentaler Zustände zunutze machen konnten. Perzeption und Kognition kamen vor Simulation: Nur wer sehen kann, kann auch träumen.[55]

[54] Das heißt nicht, daß jedem dieser Sachverhalte eine durch das System intern dargestellte Proposition bzw. ein inneres Satzanalogon entsprechen muß oder daß die entsprechenden Simulate in propositionalem Format vorliegen müssen.

[55] Gilt dies auch für Sprechen und Denken? Möglicherweise müssen wir die Manipulation diskreter Symbole erst extern erlernen (durch Operationen mit äußeren, physikalischen Symbolen wie Schriftzeichen oder sprachlichen Lauten), bevor wir sie *mental simulieren* können. Dafür sprechen neuere Überlegungen in Zusammenhang mit der Stabilität von Begriffsstrukturen und der Simulation von Sprachverarbeitung in konnektionistischen Systemen, die sich auch durch empirische Untersuchungen stützen lassen. Vgl. Rumelhart, Smolensky, McClelland & Hinton 1986, Band 2; Goschke/Koppelberg 1990: 267, Helm 1991, Kapitel 6, Johnson-Laird 1990, Bechtel/Abrahamsen 1991.

Wenn dieser Punkt deutlich geworden ist, können wir zurückkehren zu solchen Fällen mentaler Simulation, die im Zusammenhang der biologischen Evolution (speziell der menschlichen) entstanden sind und fragen: Welche Funktion kann es für ein biologisches System haben, mögliche Welten intern zu simulieren? Welche biologische Funktion könnte dadurch erfüllt werden, daß nicht-existierende Welten zu Gegenständen mentaler Simulation werden? Auf der Grundlage des eben eingeführten Begriffspaares „mentale Repräsentation" / „mentale Simulation" können wir vielleicht zu einem genaueren Verständnis bestimmter Klassen von mentalen Zuständen gelangen, indem wir „philosophisch problematische" Sonderfälle des grundlegenden Phänomens analysieren und dabei unseren rudimentären Begriffsapparat immer weiter differenzieren.

Einen Auslesevorteil dürfte es nur bedeuten, wenn das System aus der Unendlichkeit möglicher Welten diejenigen extrahieren kann, die *biologisch wahrscheinliche* Welten darstellen. Es muß also eine Heuristik besitzen, die den logischen Raum komprimiert auf zwei wesentliche Klassen von „intendierten", d. h. ausleserelevanten Welten. Zuerst einmal sind dies die wünschenswerten Welten, d. h. all jene möglichen Welten, in denen das System sich optimaler äußerer Bedingungen, vieler Nachkommen und eines hohen sozialen Status' erfreut. Diese Welten sind interessante Simulanda, wenn es um mentale Zukunftsplanung geht. Andererseits sind auch all jene möglichen und wahrscheinlichen Welten interessante Simulanda, in denen das System und seine Nachkommen zu Tode gekommen oder sonstwie in ihrem Fortpflanzungs- und Selbsterhaltungserfolg eingeschränkt sind. Solche Welten sind intendierte Simulanda bei der mentalen Risikoabschätzung von Verhaltensweisen.

Sollen mentale Simulationen erfolgreich sein, muß es demnach eine Möglichkeit geben, verschiedenen intern erzeugten Makro-Simulaten *(„möglichen phänomenalen Welten")* unterschiedliche Wahrscheinlichkeiten zuzuweisen. Wahrscheinlichkeitsabschätzung besteht in der Messung des Abstands möglicher Welten zur wirklichen Welt. *Mentale* Wahrscheinlichkeitsabschätzung kann demnach nur in der Messung des Abstands eines mentalen Makro-*Simulats* von einem mentalen Makro-*Repräsentat* bestehen. Es müssen also komplexe interne Systemzustände miteinander verglichen werden. Die Analyse und Erprobung neuronaler Netze hat uns hierfür eine präzise Begrifflichkeit geliefert: In einem konnektionistischen System kann man interne Simulate als Mengen von Subsymbolen bzw. Aktivierungsvektoren darstellen. Die Ähnlichkeit zweier Aktivierungsvektoren läßt sich mathematisch exakt durch den Winkel beschreiben, den sie im Vektorraum bilden.[56] Ohne diesen technischen Punkt weiter zu vertiefen[57], möchte ich darauf hinweisen, daß in der Notwendigkeit einer Funktion zur

[56] Vgl. zum Beispiel Churchland 1989, Helm 1990: 184.

[57] Vgl. jedoch Abschnitt 2.2.1. Gute einführende Textsammlungen sind Anderson/ Rosenfeld 1988, Nadel *et al.* 1989 oder Rumelhart, Smolensky, McClelland & Hinton 1986.

Wahrscheinlichkeitsabschätzung der Schlüssel zu einem besseren Verständnis desjenigen Aspekts von Subjektivität liegen könnte, den wir meistens unter der Überschrift „Bewußtsein" diskutieren. Es muß nämlich ein *Medium* geben, in dem diese Abschätzung stattfinden kann. Das heißt: Es muß durch Generierung eines geeigneten inneren Systemzustands ein höherstufiges Meta-Repräsentat erzeugt werden, das den „Wahrscheinlichkeitsabstand" zwischen Simulat und Repräsentat (zum Beispiel den Winkel zwischen zwei Aktivierungsvektoren) nochmals mental abbildet. Dies könnte die Funktion von Bewußtsein sein.[58] Bevor wir uns einer repräsentationalen Analyse der Bewußtheit mancher mentaler Repräsentate zuwenden, müssen wir jedoch ein klareres Verständnis eines genetisch früheren Entwicklungsstadiums phänomenaler Zustände zu erreichen versuchen. Dies ist das Ziel des nun folgenden Abschnitts.

2.1.3 Mentale Präsentation: Qualia

Zumindest *intendierte* Simulationen[59] unterscheiden sich von Repräsentationen bezüglich einer Reihe von subjektiven Eigenschaften der durch sie aktivierten phänomenalen Zustände. Wir wissen, daß unsere absichtlich eingeleiteten inneren Monologe oder unsere sexuellen Phantasien nicht in einem direkten kausalen Verhältnis zur Welt stehen, weil ihnen die perzeptuelle Komponente und damit ein Großteil des qualitativen Gehalts fehlen, der genau dieselben Bewußtseinsinhalte begleiten würde, wenn sie durch Repräsentation ausgelöst würden und nicht durch Simulation. Phantasien sind mentale *Konstrukte* – etwas Gemachtes, das nicht den vollen Erlebnisgehalt von psychischen Zuständen mit einer „Weltkomponente" trägt. Solche mentalen Prozesse sind nur innere Erlebnisse, weil sie durch interne

[58] Somit haben wir bereits ein **zweites** gutes Argument für die „teleofunktionalistische" Notwendigkeit von Bewußtheit. Von der zugrundeliegenden metarepräsentationalen Funktion muß man allerdings verlangen, daß sie alle möglichen Simulate erfassen kann.

[59] *Intendierte Simulationen* sind solche Prozesse, bei denen das psychologische Subjekt sich als den Initiator der jeweiligen Sequenz innerer Zustände erlebt. Was *intendiertes Agieren* angeht, lassen die Untersuchungen Benjamin Libets zur Willkürmotorik und zur Antedatierung somästhetischer Reize durch das Gehirn Zweifel an der Souveränität des psychologischen Subjekts gegenüber „unbewußten zerebralen Initiativen" aufkommen. Vgl. hierzu LIBET 1973, Libet et al. 1979, Libet 1983 und Libet 1985. Leider kann ich hier nicht weiter auf die Libetschen Daten und das philosophische Problem der *phänomenalen Datierung mentaler Ereignisse* im Gegensatz zur *physikalischen Datierung ihrer neurobiologischen Korrelate* eingehen. Ich glaube, daß die Bedeutung dieser empirischen Resultate für eine Theorie des Geistes – die uns ja unter anderem die Beziehung zwischen Erlebnissubjekt und Handlungssubjekt erläutern muß – bisher unterschätzt worden ist. Eine Überblick über die Problematik gibt die Diskussion von Libet 1985 in BBS (1985) **8:** 529–566. Eine Darstellung der Äquivokationen von „Bewußtsein" in der Interpretation des Materials gibt Bisiach 1988: 105ff; vgl. dazu auch die dualistische Interpretation der Daten von Eccles 1976, und Popper/Eccles 1982: 315ff oder die in eine ganz andere Richtung zielende Diskussion zwischen Frau Churchland und Benjamin Libet (Churchland 1981a, 1981b, Libet 1981).

Ereignisse verursacht werden: Ihre Genese ist in kausaler Hinsicht auf systeminterne Zustände beschränkt, was sich in einem partiellen Verlust ihres phänomenalen Gehalts niederschlägt. Diese Reduktion von phänomenalen Gehalt muß eine Theorie des Geistes erklären können. Denken wir deshalb noch einmal an sexuelle Phantasien (stellen wir also einen Typ von mentaler Simulation durch einen anderen Typ von mentaler Simulation dar): Was unterscheidet imaginierten Hautkontakt von *tatsächlichem*[60] Hautkontakt? Zunächst sind es eine Reihe von taktilen und haptischen Qualitäten wie „Wärme" oder „Zartheit", die uns in der Fantasie weitgehend fehlen. Die Entstehungsbedingungen solcher Qualitäten sind von zentralem Interesse für jede Theorie der Subjektivität; ich komme deshalb bald auf sie zurück. Außerdem fehlt jedoch dem mental simulierten Gegenstand der sexuellen Phantasie eine gewisse erlebbare Objekthaftigkeit, eine phänomenale Widerständigkeit – und zwar deshalb, weil er aus der Perspektive des psychologischen Subjekts als *gemacht* erscheint und nicht als *gegeben*.

Diese Qualität der *Gegebenheit*[61] ist eine der philosophisch interessantesten phänomenologischen Eigenschaften vieler subjektiver Bewußtseinsinhalte. Sie stellt eine der Hauptquellen für unsere dualistischen Intuitionen dar: Wie kann die *instantane* Qualität von Schmerzen oder eines Rot-Erlebnisses durch physische Prozesse erzeugt werden? Soll nämlich der Erleb-

---

[60] Achtung! Auch perzeptuell präsentierter Hautkontakt ist nicht „wirklicher" Hautkontakt im Sinne des naiven, „direkten" Realismus.

[61] Diese phänomenale Qualität bildet das unsichere Fundament vieler spekulativer und metaphysischer Theorien des Geistes. Der prominenteste empiristische Kritiker des immer wieder anzutreffenden *Myth of the Given* ist Wilfrid Sellars (Vgl. Sellars 1963; eine deutsche Übersetzung der Abschnitte XI. bis XV. ist unter dem Titel *„Der Empirismus und die Philosophie des Geistes"* in Bieri 1981: 184ff erschienen). Darstellungen finden sich in Kurthen 1988: 74–85 und Kurthen 1990, Abschnitt 2.2. Kurthen schreibt über die Stoßrichtung der Sellarsschen Kritik naiv-realistischer Interpretationen der von mir angesprochenen phänomenalen Qualität: *„Dabei richtet er sich im wesentlichen gegen diejenigen Varianten des Empirismus, die im Gegebenen als einer epistemologischen Kategorie das nichtinferentielle Fundament empirischen Wissens sahen. Aber im Grunde ist das „Gegebene", das Sellars sich als Gegner aufbaut, ein recht komplexes philosophisches Vorurteil, das sich zumindest aus den folgenden Intuitionen zusammensetzt:*

- das Gegebene hat den phänomenalen Charakter der Unmittelbarkeit, Gegebensein ist unmittelbares Gewahrhaben (awareness);
- es ist ein nichtinferentielles Wissen um phänomenale Qualitäten;
- das Vorkommen von Gegebenem im Bewußtsein setzt *keinen Lernprozeß voraus;*
- die betreffende Person hat einen privilegierten und direkten Zugang zu dem ihr Gegebenen;
- *das Entstehen des Gegebenen könnte man so illustrieren: die kategoriale Struktur der Welt prägt sich dem Bewußtsein auf wie das Muster des Siegels dem Wachs;*
- „the idea that. . .the noninferential knowledge of facts belonging to this structure constitutes the ultimate court of appeals for all factual claims about the world" (Sellars 1963: 164);
- „the idea that observation ‚strictly and properly so-called' is constituted by self-authenticating nonverbal episodes, the authority of which is transmitted to verbal and quasi-verbal performances. . ." (Sellars 1963: 169, Kurthen 1990: 21f.)

nischarakter unserer inneren Zustände im Rahmen einer Theorie mentaler Repräsentation erläutert werden, so müssen wir unter anderem eine Erklärung dafür anbieten können, wie es möglich ist, daß aus komplexen zeitlichen Prozessen auf der neurobiologischen Ebene Formen von mentalem Gehalt entstehen können, die dem introspizierenden Erlebnissubjekt als zeitlich *unmittelbar* erscheinen.[62]

Die Vertreter des metaphysischen Subjektbegriffs könnten zum Beispiel erwidern, daß etwa Schmerzerlebnisse *immer schon* als solche gegeben sind. Zwar sind sie deutlich episodisch, jedoch ist die Qualität der Schmerzhaftigkeit instantan, d. h. es gibt keine phänomenale *Genese.* Die Freunde einer anti-naturalistischen Interpretation von Qualia könnten zudem ins Feld führen, daß sie als phänomenale Zustände entweder in ihrem vollen qualitativen Gehalt gegeben oder nicht-existent sind. Dieser Gehalt kann sich ändern, möglicherweise an Intensität zu- oder abnehmen, er besitzt jedoch keine innere Geschichte. Die Qualität der Schmerzhaftigkeit selbst ist *ahistorisch.* Deswegen kann man dafür argumentieren, daß die Schmerz-Quale unmöglich das Resultat von Prozessen in Raum und Zeit sein kann. Wie könnte etwas, das unvermittelt auftaucht und verschwindet, subjektiv vor aller Reflexion oder Introspektion immer schon da ist, jemals in Begriffen nicht-psychologischer Beschreibungsebenen analysiert oder gar auf Neurobiologie reduziert werden? Mit der Gegebenheit geht außerdem ein schwer abzuweisendes introspektives Gewißheitserlebnis einher:

> Nothing, it seems, could you know more intimately than your own qualia; let the entire universe be some vast illusion, some mere figment of Descartes' evil demon, and yet what the figment is made of (for you) will be the qualia of your hallucinatory experiences. Descartes claimed to doubt everything that could be doubted, but he never doubted that his conscious experiences had qualia, the properties by which he knew or apprehended them.[63]

Es gibt noch eine Reihe anderer Gründe, die für einige Philosophen eine dualistische Theorie von Qualia nahelegen. Bevor ich auf diese Gründe eingehe, möchte ich kurz skizzieren, wie man eine naturalistische Perspektive auf die fragliche Klasse mentaler Zustände einnehmen kann. Man kann dies tun, indem man einen bestimmten Typ von innerem Zustand auszuzeichnen versucht und ihn auf seine biologische Funktion hin untersucht. Wenn die biologische Funktionalität gewisser Typen mentalen Gehalts deutlich wird, verlieren metaphysische Interpretationen automatisch an Attraktivität und der Weg zu einer empirisch gehaltvollen Präzisierung wird frei. Darum werde ich nun zusammen mit einigen Erläuterungen den Begriff des mentalen Präsentats einführen, um dann einen kurzen Blick auf einige mit ihm verknüpfte philosophische und empirische Probleme zu

[62] Die neuropsychologischen Grundlagen der eben angesprochenen phänomenalen Qualität der „Gegebenheit" in bezug auf subjektives Zeiterleben hat Ernst Pöppel untersucht. Vgl. Pöppel 1985, 1989.

[63] Vgl. Dennett 1988: 42.

werfen. Am Ende dieses Abschnitts werde ich eine vorläufige und recht spekulative Hypothese bezüglich Qualia aus der Perspektive einer naturalistischen Theorie mentaler Repräsentation formulieren.

Wie wir bereits gesehen haben, war die Geschichte subjektiven Bewußtseins auf unserem Planeten eine *biologische* Geschichte. Eine der frühesten und grundlegendsten Aufgaben von Bewußtsein wird es gewesen sein, zu präsentieren: Interne Ereignisse zu erzeugen, die zuverlässig die *Gegenwart* äußerer Eigenschaften der Welt anzeigen. Solche Zustände könnten sich von reinen Reflexbögen dadurch unterschieden haben, daß sie ein stabiles internes Präsentat erzeugten. Insofern solche inneren Zustände bereits das Potential besitzen, vorübergehend zu Inhalten subjektiven Bewußtseins zu werden, kann man sie auch als mentale Präsentate bezeichnen.[64] Mentale Präsentate besitzen zwei Eigenschaften, die für das philosophische Qualia-Problem von zentraler Bedeutung sind.

Erstens: Sie sind nicht unabhängig vom Strom des Inputs aktivierbar. Mentale Präsentate haben *Signalcharakter*, d. h. sie signalisieren *für* ein System die aktuelle Präsenz eines Präsentandums. Mentale Präsentation ist also eine rudimentäre Form von mentaler Repräsentation. Schmerzerlebnisse oder eine türkise Farbwahrnehmung zum Beispiel können wir nicht unabhängig von einer Signalquelle in uns erzeugen. Denn es gibt in nichtpathologischen Zuständen keine Möglichkeit einer phänomenalen Simulation von Qualia: Mentale Präsentate unterscheiden sich von mentalen Repräsentaten dadurch, daß sie nicht simulationsfähig sind. Wir können ein Schmerzerlebnis oder ein türkises Seherlebnis nicht einfach mit geschlossenen Augen in uns erzeugen. Das bedeutet, daß sie im allgemeinen in Abwesenheit der üblichen kausalen Antezedentien nicht aktivierbar sind.[65] Weil Präsentate nicht die kausalen Relationen von Umweltelementen untereinander abbilden, sondern die pure aktuelle Präsenz einer Reizquelle zum Inhalt haben, hat es auch den Anschein, als entzögen sie sich einer funktionalen Analyse.[66] In Qualia erleben wir ein reines, nicht-relationales[67] Präsentieren. Aber selbstverständlich tragen Qualia *Information*[68], und zwar

[64] Meine vorsichtige Skizze des Begriffs der „mentalen Präsentation" sollte nicht als eine direkte Anknüpfung an den Begriff der „Gegenwärtigung" bei Husserl und Heidegger oder andere frühere Begriffe der Präsentation etwa bei Meinong, Spencer oder Bergson gelesen werden. Vgl. insbesondere Metzinger 1998.

[65] Ausnahmen sind hier zum Beispiel *Eidetiker*, die vollständige Farberlebnisse bei geschlossenen Augen haben können. Die Erzeugung interner Signalquellen (zum Beispiel durch Visualisierungsübungen oder durch pharmakologische Enthemmung gewisser Hirnbereiche) führt zu ähnlichen subjektiven Phänomenen. Diese Phänomene widersprechen nicht der hier vertretenen Theorie über mentale Präsentate, weil in diesen Fällen der notwendige Input intern generiert wird.

[66] Nemirov 1979 bietet eine tiefgehende Analyse des Problems subjektiver Erlebnisqualitäten aus der Perspektive eine funktionalistischen Theorie des Geistes.

[67] Gegen diese Auffassung argumentiert Dennett 1988.

[68] Es kann also auch *Fehlpräsentationen* geben, wenn nämlich das dem Präsentat zugehörige mentale Modell (vgl. Abschnitt 2.2) die kausale Relation, in der das Präsentat zu einer Signalquelle steht, falsch interpretiert: Die Farben eines geträumten Regenbogens haben ihre

über die gegenwärtige Präsenz einer Reizquelle. Bei *Re*-Präsentationen (wie beim *Nachdenken* über Schmerzen oder beim *Sich-Erinnern* an ein grünes Nachbild) fehlt dieser Aspekt wegen seiner Input-Abhängigkeit zumeist. Interessant ist, daß Präsentate nur selten isoliert auftreten – etwa wenn die „Hardware" des visuellen Cortex „nachglüht" und Nachbilder in Komplementärfarben entstehen läßt.[69] Im Normalfall sind der Präsentationsaspekt und der Repräsentationsaspekt interner Zustände bruchlos integriert.

Mentale Präsentation: $\mathbf{M}_{\text{Prä}}$ (S,X,Y)
- X präsentiert Y *für* S.
- X ist ein Präsentat; ein nicht simulationsfähiger interner Systemzustand, der nur bei konstantem Input aufrechterhalten werden kann.
- $\mathbf{M}_{\text{Prä}}$ ist antisymmetrisch.

Zweitens: Unsere cartesianischen Intuitionen bezüglich der Transparenz und Unmittelbarkeit phänomenalen Bewußtseins haben neben dem puren Präsentationsaspekt mancher mentalen Zustände ihre Wurzeln in mindestens einer weiteren Eigenart dieser Zustände. Viele von ihnen besitzen eine instantane Qualität. Was heißt das? Es bedeutet, daß zu ihrem subjektiven Erlebnischarakter auch eine *Unmittelbarkeit im zeitlichen Sinne* gehört. Wenn die mit ihnen verknüpften mentalen Gehalte subjektiv gegeben sind, dann ist auch der pure Präsentationsaspekt immer schon gegeben. Der rote Apfel auf dem Tisch ist immer schon rot, der qualitativ-präsentative Aspekt der Röte ist unter Standardbedingungen immer schon Teil des Gesamterlebnisses. Interessanterweise ist er das aber ohne in Relation zu anderen Eigenschaften des Apfels (räumliche Ausdehnung, Form, Gewicht) oder weiteren Elementen des phänomenalen Feldes zu stehen.[70]

Ursache in der Enthemmung gewisser Regionen des Hirnstamms und nicht in retinalen Erregungsmustern. Der pure Präsenzaspekt dagegen kann nicht fehlpräsentiert werden, denn immer wenn ein bestimmtes Farbquale auftritt, sind ganz bestimmte Datenstrukturen in speziellen Formaten in den entsprechenden Hirnregionen aktiviert worden. Mentale Präsentate tragen also zuverlässige Information über einen gewissen Systemzustand. Die sie begleitenden mentalen Modelle können dagegen fehlgehen, was die Darstellung der kausalen Genese dieses Systemzustands angeht. – Das Jacksonsche *Knowledge-Argument* ist häufig auch als Hypothese über phänomenale Information interpretiert und diskutiert worden. Vgl. Dennetts Bemerkungen zu Peter Bieris „PIPS-Hypothese" (Dennett 1988: 71f) und Lewis 1988.

[69] Wenn Sie in ein quadratisches rotes Blitzlicht schauen und dann die Augen schließen, erleben Sie nach einiger Zeit nur noch einen diffusen grünen Farbfleck. Einige Zeit nachdem das relationale mentale Modell des Blitzlichts verschwunden ist wird das (komplementäre) mentale Präsentat noch aufrechterhalten. Genaugenommen handelt es sich hierbei um eine Ermüdungserscheinung (darum bewegen sich Nachbilder auch mit den Augen). Vgl. Hardin 1988: 93.

[70] Diese Beobachtung kann man bezweifeln, indem man Sonderfälle oder „geübte Introspektoren" untersucht; vgl. Dennett 1988, Hardin 1988. Wenn es stimmt, daß Qualia durch Metarepräsentation anhand von Formaten individuierte Systemzustände sind, dann könnte

Die Kombination der Unabhängigkeit von anderen relationalen Inhalten und des erlebnismäßigen Immer-schon-Gegebenseins kann zu einer transzendentalen Interpretation von Qualia verleiten: Dann erscheinen Qualia auch aus theoretischer Sicht als etwas vor aller (inneren) Erfahrung bereits Feststehendes, von ihr Unabhängiges und deshalb Außerweltliches. Wenn man die intuitive Evidenz introspektiver Erlebnisse nicht erkenntniskritisch hinterfragt, gelangt man auf diese Weise fast zwangsläufig zur Postulierung transmundaner Objekte.

Gegen die Freunde des metaphysischen Subjekts, die diese Eigenart mancher mentaler Zustände theoretisch ausbeuten möchten, läßt sich aus naturalistischer Perspektive einwenden: Es gibt gute neuroinformatische Alternativerklärungen. Wenn die zugrundeliegenden Prozesse neuronaler Informationsverarbeitung schlicht zu schnell sind, als daß sie zu Repräsentanda von Introspektion werden könnten[71], dann wird den so erzeugten Präsentaten auf der phänomenalen Ebene bewußten Erlebens die Qualität des „Gewordenseins" oder „intern Konstruiertseins" fehlen. Dieses Phänomen ist keineswegs ungewöhnlich, es tritt häufig auch in Zusammenhang mit höheren kognitiven Leistungen auf: Die Lösung eines Problems oder der Ort, an dem wir einen attraktiven Mann oder eine attraktive Frau zum letzten Mal gesehen haben, kann uns als psychologischen Subjekten gewissermaßen „in unseren inneren Erlebnisraum" *einfallen* (nämlich unerwarteterweise, sozusagen *nachdem* wir das Vertrauen auf die in unserem Rücken „für uns" arbeitende neuronale Maschinerie bereits verloren haben). In solchen Fällen mag es eher so sein, daß die fehlende Metarepräsentation nicht stattfindet, weil sie in der zu einem nicht unbeträchtlichen Teil genetisch determinierten funktionalen Architektur unseres Gehirns nicht „vorgesehen" ist. Wenn mentale Repräsentate und Präsentate die Resultate interner Informationsverarbeitungsvorgänge sind, dann bedeutet das nämlich nicht automatisch, daß die Zeitlichkeit der zugrundeliegenden Vorgänge durch ihren Gehalt noch einmal dargestellt wird. Falls meine Hypothese richtig ist, daß die zuletzt genannten Zustände primär Signalcharakter besitzen, dann ist dies sogar teleofunktionalistisch leicht einsehbar. Wenn Qualia funktionale Zustände sind, deren Funktion im Signalisieren der puren Präsenz

---

man sagen: Die Formate sind *plastisch* (d. h. differenzierbar) und durch eine Erhöhung der „introspektiven" bzw. metarepräsentationalen Fähigkeiten kann das System Beziehungen zwischen ihnen entdecken (zum Beispiel die Wahrscheinlichkeit des gemeinsamen Auftretens, Exklusivität usw.) .

[71] Ernst Pöppel hat in seinen Untersuchungen zur zeitlichen Quantelung der Informationsverarbeitung im Gehirn philosophisch interessante Ergebnisse bezüglich der neurobiologischen Rahmenbedingungen, die das „Gegenwartsfenster" eines Organismus determinieren, zutage gefördert. Hier zeigt sich, wie das „phänomenale Jetzt" eines Systems in seiner Ausdehnung durch die spezifische Weise bestimmt wird, in der der Informationsfluß im Gehirn „getaktet" ist. Eine solche Taktung der Informationsverarbeitung ist in modularen Systemen notwendig, weil unterschiedliche Transduktionsgeschwindigkeiten ihrer Subsysteme (zum Beispiel des visuellen und des auditorischen Cortex) den „Gleichzeitigkeitshorizont" externer Ereignisse festlegen, die vom System intern dargestellt werden. Vgl. Pöppel 1985, 1989.

eines Sachverhalts in der Welt *für* das jeweilige System besteht, dann ist es natürlich sinnvoll, daß diese Zustände so schnell wie möglich erzeugt werden und die relationalen Eigenschaften sowie die Prozessualität ihrer kausalen Antezedentien nicht darstellen. Schmerzen *müssen* schnell sein, um ihre biologische Funktion – die Auslösung erfolgreichen Vermeidungsverhaltens – zu erfüllen. In jedem Fall gibt es gute und überzeugende neuroinformatische Ansätze zur Erklärung der Eigenschaften derjenigen internen Datenstrukturen, die ich als „mentale Präsentate" begrifflich etwas genauer zu fassen versucht habe.

Die innerpsychisch unhintergehbare Gegebenheit und Instantanität vieler mentaler Zustände scheinen also einer naturalistischen Analyse durchaus zugänglich zu sein. Wie steht es jedoch mit den jeweiligen Qualitäten, die uns durch mentale Präsentate zeitlich unmittelbar und nicht-derivativ gegeben sind? Ein Schmerzerlebnis oder eine Rotwahrnehmung präsentieren dem psychologischen Subjekt scheinbar eine pure Qualität, eine *phänomenale Essenz*. Diese Qualität hat einen monadischen Charakter, sie scheint der private *Kern* des Erlebnisses zu sein. Sie ist außerdem inkommunikabel: Wir können einem Blinden nicht erklären, was Röte ist. Fred[72] kann uns nicht begreiflich machen, *welche* Art von innerem Erlebnis es ist, die ihm gestattet, zwei Arten von Tomaten zu unterscheiden, die für uns einfach nur rot sind. Außerdem weist die subjektive Qualität eines Präsentats *prima introspectione* keinerlei Beziehungen zu anderen Elementen der phänomenalen Ebene auf, sie kann nicht durch andere Elemente analysiert werden (Wir können Mary nicht weiterhelfen, indem wir ihr Röte über Grautöne ihres achromatischen Bildschirms erläutern[73] oder durch transmodale Analogien[74] wie *„Rot ist wie der Ton einer Trompete"*). Diese – von den Verteidigern des metaphysischen Subjektbegriffs immer wieder ins

[72] Vgl. Jackson 1982 und Kapitel 4.

[73] Man kann allerdings subjektive Farberlebnisse durch schwarzweiße Signale auslösen, die in der richtigen Weise pulsieren. Beispiele für solche Stimuli sind eine mit 6–8 Hz rotierende (schwarz-weisse) *Benham*–Scheibe oder auch die schwachen, ungesättigten Farben, die man im optischen Rauschen des nicht besetzten Kanals eines Schwarzweißfernsehers sehen kann. J. F. Butterfield hat gezeigt, wie man mit einem auf die richtige Weise animierten Schwarzweißbild *„subjektives Farbfernsehen"* erzeugen kann; vgl. Butterfield 1968, 1970. Den Hinweis auf diese Möglichkeit (mit der man die Jacksonsche Mary sogar bei Einhaltung der Regeln des Gedankenexperiments mit visuellen Qualia beglücken könnte) verdanke ich dem ausgezeichneten Buch von Larry Hardin. Vgl. Hardin 1988: 72, 1990; Jackson 1982.

[74] Warum ist Blau die *Farbe der Ferne*, die Farbe der Weite und der Sehnsucht? Weil es eine gute visuell-kinetische Analogie gibt, nach der menschliche Wesen Farberlebnisse subjektiv als *„sich annähernd"* oder *„sich zurückziehend"* kategorisieren können: *„If the lens of the eye accomodates so that an image formed by middlewave light is focused on the retina, a shortwave image will focus in front of the retina, and a longwave image will focus behind it. Shortwave light thus focuses like light from a more distant object, and longwave light focuses like light from a closer object. To focus on a blue patch we must therefore accomodate our lenses as we would when we focus on a distant object, while focusing on a red patch requires that we accomodate as we do for nearby objects. It is sometimes claimed that afferent messages from the relevant eye muscles are among the clues we use to establish the respective distances of objects."* Hardin 1988: 129.

Feld geführte – putative Irreduzibilität von Qualia hat monistischen Philosophen seit jeher Sorge bereitet. Einige von ihnen haben Qualia ignoriert[75], andere haben versucht, sie zu eliminieren[76], zu Epiphänomenen zu degradieren[77] oder schlicht ihre Existenz überhaupt bestritten[78] und ich werde weiter unten einige ihrer Argumente diskutieren. Wichtig scheint mir dabei jedoch, daß eine naturalistische Theorie des Geistes den qualitativen Gehalt mentaler Zustände ernst nimmt. Gegenüber den Theoretikern vergangener Jahrhunderte besitzen wir den Vorteil, daß wir ein wesentlich besseres und sich ständig vergrößerndes Wissen über die physischen Entstehungsbedingungen phänomenaler Qualitäten besitzen. Es wäre falsch, auf diesen Vorteil mit physikalistischer Arroganz gegenüber dem Problem zu reagieren.

Was also ist *allen* Roterlebnissen gemeinsam? Ihnen ist gemeinsam, daß sie mit der Erregung bestimmter, eng umgrenzter Areale innerhalb eines bestimmten Moduls unseres Hirns[79] einhergehen und daß sie durch Stimulation dieser Areale zuverlässig ausgelöst werden können. Roterlebnisse sind mentale Präsentate, die in ihrem Auftreten eng korreliert sind mit der Erzeugung von Datenstrukturen eines gewissen Typs, nämlich mit Datenstrukturen, die immer einem bestimmten Modul bzw. dem *Subsystem* eines solchen Moduls entstammen. Module sind funktionale Untereinheiten informationsverarbeitender Systeme, die – um die Terminologie von Jerry Fodor zu übernehmen – informationell eingekapselt[80] sind. Das heißt: Ihr interner Informationsfluß ist weitgehend von dem anderer Prozessoren abgeschottet. Das Gehirn als ein natürliches Repräsentationssystem mit einer biologischen Geschichte von vielen Millionen Jahren unterscheidet sich aber von den künstlichen Repräsentationssystemen, die wir bis heute kennen, in vielen Punkten. Ein theoretisch sehr interessanter Unterschied zwischen biologischen und artifiziellen Maschinen scheint zu sein, daß die Formate, in denen Datenstrukturen erzeugt und verarbeitet werden, im Gehirn nicht durch einen Programmierer festgelegt werden.

---

[75] Dies gilt für die frühe Phase des Maschinenfunktionalismus. Vgl. Putnam 1975, dazu auch Nemirow 1979.

[76] Vgl. Churchland 1981b, 1985b.

[77] Vgl. Jackson 1982.

[78] Vgl. Dennett 1988.

[79] Es kann auch sein, daß sich die Modularität nicht direkt anatomisch ausdrückt, sondern in Form *lokaler Minima* in der Energielandschaft eines neuronalen Netzes. Man könnte in solchen Fällen zwischen Makro- und Mikroformaten unterscheiden: das visuelle oder auditorische Format mentaler Repräsentate beruht auf Eigenheiten eines neuroanatomisch abgrenzbaren *Moduls*, wogegen *einzelne* Farben oder Klänge durch lokale, stabile Aktivierungszustände innerhalb eines Moduls (in die das System sich „relaxieren" kann) repräsentiert werden. Der klassische Text zur Modularität phänomenalen Bewußtseins ist Fodor 1983; Ray Jackendoff hat auf die Bedeutung modalitätenspezifischer Repräsentationshierarchien (und ihrer Vereinigung im dreidimensionalen Modell des visuellen Apparats) hingewiesen sowie die Fodorsche Modularitätsthese verfeinert und weiterentwickelt. Vgl. Jackendoff 1987, Kapitel 12. Vgl. auch Gazzaniga 1988.

[80] Vgl. Fodor 1983, III.5, dazu auch Jackendoff 1987, Kapitel 12.

Auf einem Computer können wir festlegen, in welchem Format wir Information repräsentieren und ausgeben lassen. Zum Beispiel können wir Information in Form von Bildern und in Form von kunstsprachlichen Satzanaloga (Maschinensprachen) intern repräsentieren und verarbeiten oder auch extern darstellen lassen. Wir können die Maschine Piktogramme oder Sätze, Graphiken oder Zeichenketten konstruieren lassen. Unvermeidlicherweise sind über die Bedeutung piktorialer und propositionaler Repräsentationsformate für eine Theorie des Geistes auch bereits erbitterte Fachdiskussionen entbrannt.[81] Vielleicht haben wir gerade erst zu verstehen begonnen, wie wichtig der Begriff des „Formats" für ein naturalistisches Verständnis phänomenalen Bewußtseins ist. Dies ist nur zu verständlich, da Gehirne wahrscheinlich eine wesentlich größere Anzahl von Formaten erzeugen und benutzen als unsere heutigen Computer. Das liegt daran, daß sie auf der Ebene ihrer physischen Realisierung weit stärker modularisiert sind. In anderen Worten: Biologische Informationsverarbeitungssysteme wie das menschliche Gehirn haben bereits auf der „Hardware-Ebene" wesentlich mehr funktionale Subsysteme.[82] Man kann nun annehmen, daß jedes dieser Subsysteme, soweit es vom restlichen Informationsfluß abgeschottet ist, mit eigenen *internen Formaten* arbeitet. Wenn diese Vermutung richtig ist, dann werden eine Vielzahl von Modulen ihren Output auch in einer Vielzahl von Formaten weitergeben. Diese internen Formate könnten abstrakte Eigenschaften neuronaler Erregungsmuster sein, die durch ihre *Position in Vektorräumen* genauer beschrieben werden können.

Die Formate einzelner Datenstrukturen haben nun interessante Eigenschaften mit dem qualitativen „Kernaspekt" mentaler Präsentate gemein: Sie sind inkompatibel mit anderen Formaten und weisen – so sie nicht in einem noch höheren Format integriert werden – untereinander keinerlei informatische Beziehung auf. Ich will hier keine generelle Identitätsthese[83] bezüglich Typen von mentalen Präsentaten und Typen neuronal realisierter Datenstrukturen vertreten (etwa zwischen dem Farbaspekt visueller Erlebnisse und bestimmten Partitionen von Vektorräumen, durch die Subsysteme des visuellen Cortex beschrieben werden können). Trotzdem könnte die phänomenale Qualität der Röte einfach ein bestimmtes Repräsentationsformat sein, das von gewissen Subsystemen des menschlichen Gehirns verwendet wird. Die phänomenale Atomizität und Irreduzibilität auf andere phänomenale Elemente könnte **erstens** aus der Inkompatibilität der jeweiligen Formate resultieren und **zweitens** daraus, daß der Vorgang, durch den die fraglichen Repräsentate aus niedrigstufigeren Repräsentaten mit eigenen Formaten erzeugt werden, selbst *kein* Repräsentandum menta-

[81] Siehe Fußnote 34.

[82] Letztlich bedeutet dies, daß auch die philosophische *top-down*-Analyse des Mentalen Schritt für Schritt zu einem *Mikrofunktionalismus* übergehen wird. Vgl. etwa Clark 1989: 34.

[83] Zur Logik und Schwierigkeiten der Identitätstheorie vgl. Beckermann 1990, Bieri 1981: 31ff, Metzinger 1985: 16ff & 210ff.

ler Meta-Repräsentation ist. Und im Fall des Farbensehens kann man sogar sagen, daß der qualitative Aspekt visueller Präsentate dem psychologischen Subjekt eine wichtige Information anbietet:[84] Nämlich durch *welches Subsystem* eines Sinnesmoduls der fragliche Bewußtseinsinhalt geliefert wird. In diesem Sinne kann man sagen: Der qualitative Aspekt mentaler Präsentate trägt Information, und zwar über seine *physische Genese.* Die Neurowissenschaftlerin Mary kann, wenn ihr Gehirn zum ersten Mal Farbpräsentate erzeugt, sofort auch Aussagen über ihre internen physischen Eigenschaften machen.[85]

Der vorwissenschaftliche Begriff von Qualia ist, wie zum Beispiel Dennett überzeugend gezeigt hat[86], voller Widersprüche und eignet sich nicht für ein adäquates Verständnis unserer phänomenalen Zustände. Die philosophische Analyse zeigt ihrerseits die Unmöglichkeit, das Konzept mit einem präzisen begrifflichen Gehalt zu versehen und dabei unsere essentialistischen Intuitionen zu retten. Man kann nun die Elimination der diffusen theoretischen Entität „Qualia" favorisieren (wie es Daniel Dennett und andere getan haben) oder man kann versuchen, das Problem weiterhin ernstzunehmen, indem man den vortheoretischen Begriff klärt und vorsichtig in einen technischen Term überführt – zum Beispiel vor dem Hintergrund einer am Modell der Informationsverarbeitung orientierten naturalistischen Theorie mentaler Repräsentation. Dies werde ich nun – in zugegebenermaßen sehr spekulativer Weise – zu tun versuchen, in dem ich die folgende provisorische Hypothese aufstelle:

> **(AIT):** Qualia sind **Analog-Indikatoren**, die für ein System die aktuelle Gegenwart eines Präsentandums intern signalisieren. Das, was alle introspektiv erlebbaren mentalen Präsentate (die durch den Vorgang erzeugten aktiven Datenstrukturen) eines phänomenalen Typs (eines Quale) miteinander gemein haben, ist ihr *Format*.

Hierzu nun einige Erläuterungen. Was ist ein Indikator? Betrachten wir dazu kurz die entsprechende Funktion in natürlichen Sprachen. Indikato-

---

[84] Der evolutionäre Vorteil des Farbensehens besteht in der Ex*traktion einer zusätzlichen Umwelteigenschaft*: In Feldern gleicher Luminosität werden stärkere Kontraste erzeugt, weil nun in gewissem Ausmaß neben der energetischen Intensität auch noch die *Frequenz* des Reizsignals intern dargestellt werden kann.

[85] Gibt es *reine* Qualia, die keinerlei Präsentationsfunktion *für* das System haben? Emotionale Artefakte könnten Kandidaten für solche Zustände sein: Angenommen, wir injizieren Ihnen im Rahmen eines wissenschaftlichen Experiments eine psychoaktive Substanz, die eine völlig grundlose Euphorie in Ihnen auslöst. Plötzlich tritt nun das Quale einer reinen, grundlosen Freude in Ihrem inneren Erlebnisraum auf. Zumindest für wissenschaftlich gebildete Versuchspersonen wie Frank Jacksons *Mary* trägt dieses Quale Information: Nämlich über physische Eigenschaften eines Subsystems (zum Beispiel des limbischen Systems) ihrer *selbst*.

[86] Vgl. Dennett 1988.

ren oder indexikalischen Ausdrücken werden wir im vierten Kapitel in Zusammenhang mit der semantischen Diagnostizierung der Nagelschen Thesen erneut begegnen: „Ich", „Hier", „Dieses da", „Jetzt" sind Beispiele für solche Ausdrücke. Ihre Referenz hängt von dem räumlichen, zeitlichen oder psychischen Kontext und der Position des Sprechers in diesem Kontext ab. Sie helfen dem Sprecher, sich zu *orientieren* und zu *lokalisieren*. In gehaltvollen Aussagen können Indikatoren in ihrer Bezugnahme fehlgehen, darum möchte ich sie als *digitale Indikatoren* bezeichnen – sie erzeugen Wahrheit und Falschheit.

Analog-Indikatoren – wie zum Beispiel die von mir beschriebenen mentalen Präsentate – melden dagegen durch einen systeminternen Zustand die pure Präsenz eines Reizes. Als Datensätze sind sie unter der Hinsicht ihres Gehalts nicht relational strukturiert, aber sie besitzen relationale Eigenschaften (zum Beispiel bezüglich anderer Systemzustände und ihrer physikalischen Aktivierungsbedingungen). Wenn wir den Gehalt eines visuellen Präsentats oder einer Schmerz-Quale sprachlich wiedergeben wollten, müßten wir Digital-Indikatoren verwenden, zum Beispiel, indem wir sagen: *„Rot – Jetzt – Hier!"* oder *„Hier – Jetzt – Zahnschmerzen!"*.[87] Das verdeutlicht aus externer Perspektive die interne Funktion von Indikatoren. Warum aber *Analog*-Indikatoren? Weil mentale Präsentate einen Intensitätsparameter besitzen: Rotwahrnehmungen, subjektive Schmerzerlebnisse oder Emotionen wie Eifersucht und Dankbarkeit können innerhalb eines bestimmten Bereichs die *Intensität* oder *Signalstärke* des Präsentandums (also Information über eine physikalische Eigenschaft desselben) für den Organismus intern darstellen. Sie erzeugen aber nicht Wahrheit oder Falschheit, sondern bloß Ähnlichkeit.

An diesem Punkt wird man unweigerlich mit dem folgenden anti-naturalistischen Einwand konfrontiert: *„Nun gut, das ist ja alles gut und schön – aber es sagt uns nicht das, was wir doch immer wissen wollten: Ist die subjektive Taxonomie mentaler Zustände abbildbar auf die entstehende neurowissenschaftliche Taxonomie? Was ist denn nun am Ende die Qualität der* ***Röte*** *und die* ***Schmerzhaftigkeit*** *von Schmerzen?"* Die hypothetische Antwort lautet: Ihren Rotwahrnehmungen und Schmerzerlebnissen liegen bestimmte Datenstrukturen zugrunde, die von komplizierten neuronalen Mechanismen erzeugt werden (durch einen Vorgang, den wir auf einer höheren Beschreibungsebene als mentale Präsentation analysieren können) und die physikalisch durch bestimmte neuronale Erregungsmuster reali-

[87] Wenn man einen mentalen Zustand bei seinem ersten Auftreten nicht sofort identifizieren oder wiedererkennen kann, bedeutet dies nicht unbedingt, daß man unter einem Mangel an Faktenwissen leidet, sondern nur, daß man bestimmte Begriffe nicht erfolgreich *anwenden* kann. Gegen die diffusen Nagelschen und Jacksonschen Begriffe von „innerem Wissen" hat Janet Levin geltend gemacht, daß die Verbindung zwischen Wissen bezüglich mentaler Zustände und gewissen diskriminatorischen Fähigkeiten keine logisch notwendige ist, sondern daß wir in den fraglichen Fällen schlicht mit *epistemischen Defiziten* konfrontiert sind. Bezüglich dieser Argumentationslinie mit Blick auf das Verhältnis zwischen der objektiven und subjektiven Individuierung phänomenaler Zustände vgl. Levin 1986.

siert werden. Als Datenstrukturen besitzen sie eine *abstrakte Eigenschaft*: Das Format, in dem sie vorliegen. Die These ist, daß es genau dieses Format ist, das wir durch Metarepräsentation erfassen, wenn wir mentale Präsentate *introspektiv kategorisieren*. Daß das Format das entscheidende Merkmal sein könnte, sieht man schon daran, daß es genau der qualitative Aspekt ist, der bei Darstellung von Präsentaten in anderen Repräsentationsmedien mit anderen Formaten verlorengeht: In einem propositionalen Repräsentat nach dem Muster *„Hier – Jetzt – Zahnschmerzen!"* geht nämlich genau das Format der als „Zahnschmerz" kategorisierten Datenstruktur verloren. Der Indikator-Aspekt dagegen bleibt bestehen: Aus der *„Hier – Jetzt"*-Komponente der sprachlichen Äußerung geht immer noch eindeutig hervor, daß *in diesem System* und *gerade jetzt* eine bestimmte Datenstruktur aktiviert wird. Wenn wir annehmen, daß es sich dabei um mentales Sprechen handelt, wird der Punkt noch deutlicher: Der indexikalische Gedanke *„Hier – Jetzt – Zahnschmerzen!"* trägt (in „quasi-propositionalem" Format) dieselbe Information, aber auf ganz andere Weise, weil er ein relationales mentales Modell einer Kette diskreter Symbole ist.[88] Sein qualitativer Gehalt ist dementsprechend ein anderer. Worauf wir uns also bei Selbstzuschreibungen von Qualia beziehen, sind abstrakte Eigenschaften in uns aktivierter Datenstrukturen – eben ihr Format.

Das philosophische Problem, auf das diese Hypothese zu antworten versucht, besteht darin, daß es eindeutig phänomenale Familien[89] (Klänge, Farben, Gerüche usw.) gibt, die subjektive Erfahrung selbst aber zunächst keinerlei Hinweise darauf enthält, *wodurch* sie zu Familien werden. Unsere Strategie muß es sein, *objektive Ähnlichkeitsklassen* (etwa von Datenstrukturen mit gleichen abstrakten Eigenschaften) durch direkte Relationen mit solchen Familien von phänomenalen Zuständen zu verknüpfen. Derzeit ist es noch zu früh, um genauere Aussagen in dieser Richtung machen zu können. Es besteht aber guter Grund zu der Hoffnung, daß sich zukünftig enge Korrelationen zwischen der wissenschaftlichen Kategorisierung der vielen vom Gehirn angewandten Formate und den während ihrer Aktivierung von Versuchspersonen gegebenen introspektiven phänomenalen Gehaltszuschreibungen herausstellen werden. Dann werden wir vielleicht auch besser verstehen, wie die unterschiedliche Stärke transmodaler Analogien aus der relativen Ähnlichkeit und biologischen Geschichte der zugrundeliegenden neuralen Substrate resultiert.

> If qualitative similarities and differences among sensory states amount, in the final analysis, to similarities and differences in sensory coding, we might expect not only that there be similarities and differences across sensory modalities because of general similarities of neural processing, but that modalities whose physical substrates are more similar would show greater phenomenal similarities than those whose substrates are less similar. There is some indication that this is the case. Evolutionarily and embryologically, hearing and touch are

[88] Vgl. Abschnitt 2.2.
[89] Vgl. Hardin 1988: 132.

> much closer to each other than hearing and sight. The most powerful hearing-sight analogical property is „brightness", which in sound is best represented as the product of pitch and loudness (. . .). The correlation between pitch and hue is much weaker, and not in the direction that stimulus dimensions, particularly wavelength, would have suggested. Furthermore, no one (not even synesthetes) would confuse a sight with a sound. With touch and hearing, the situation is quite different. The vibrational sensitivity of the skin is sufficiently analogous to the vibrational sensitivity of the ear for the related sensation to be referred to as vibratory „pitch". Perceived size and spatial location properties are closely similar. At very low frequencies, tactual feeling and hearing become so similar that in the 20 Hz region one is readily confused with the other. And when small puffs of air on the forhead are used to facilitate the subjective location of clicks that are fed to a subject through earphones, the puff is sometimes taken to be the click itself.[90]

Den anti-naturalistischen Philosophen wird das nicht zufriedenstellen. Er wird vielleicht etwas verärgert sagen: *„Ihren kindlichen szientistischen Optimismus bei der Kolonisierung unserer inneren Natur in allen Ehren – aber merken sie nicht, wie unplausibel diese Strategie ist? Das Verblüffende an Qualia ist doch gerade ihre* ***Konkretheit!*** *Wie könnte die phänomenale Konkretheit der Röte oder der Schmerzhaftigkeit mit einer* ***abstrakten Eigenschaft*** *irgendwelcher neuronalen Datenstrukturen identisch sein?"* Die Antwort: Das bewußte psychische Erlebnis *konkreter* Röte oder *konkreter* Schmerzhaftigkeit entsteht dadurch, daß das Format, also die abstrakte Eigenschaft eines als Datenstruktur betrachteten mentale Präsentats, noch einmal auf die ganz spezifische Art und Weise der Metamodellierung durch unser biologisches Gehirn erfaßt wird, bevor es dadurch in unseren bewußten Erlebnisraum eintritt.[91] Qualia sind also kein letztlich physikalisches Phänomen, sondern ein *repräsentationales*: Was wir als „Quale" erleben, sind nicht neuronale Prozesse selbst, sondern die durch eine Metarepräsentationsfunktion abgebildeten abstrakten Eigenschaften einer durch diese Prozesse erzeugten Datenstruktur. Die Darstellung derselben Eigenschaft durch intersubjektive Repräsentationssysteme – zum Beispiel durch wissenschaftliche und philosophische Diskurse – ist etwas ganz anderes.

Natürlich ist **AIT** eine sehr spekulative und allgemein gehaltene Hypothese, die durch präzise empirische Daten angereichert werden muß. Selbst wenn sie in die richtige Richtung weist, wird es uns immer noch schwerfallen, uns vorzustellen, wie wir abstrakte Eigenschaften von Hirnprozessen als Konkreta erleben können[92] – weil dies unsere Fähigkeiten zur mentalen

---

[90] Vgl. Hardin 1988: 133.

[91] Vgl. dazu den nächsten Abschnitt.

[92] Interessanterweise gibt es in seltenen Fällen *idiots savants*, die – obwohl nach den üblichen Kriterien als geistig behindert einzustufen – hochabstrakte Zusammenhänge quasi-ästhetisch oder qualitativ zu erfassen scheinen. Ein Beispiel sind die von Oliver Sacks geschilderten Zwillinge, die einen Intelligenzquotienten von sechzig besitzen und die einfachsten arithmetischen Grundoperationen nicht begreifen können. Trotzdem unterhalten sie sich mit

Simulation übersteigt. Was den kontraintuitiven Rest angeht, hoffe ich auf die historische Plastizität unserer Intuitionen[93] und halte es mit Daniel Dennett: *„Any theory that makes progress is bound to be initially counterintuitive.“*[94]

Eine repräsentationale (beziehungsweise: *präsentationale*) Theorie der Qualia muß eine Reihe erstaunlicher neuerer Erkenntnisse über den menschlichen Geist erklären können. Zum Beispiel muß sie erklären können, wie es durch einen ingeniösen experimentellen Aufbau möglich ist, einem Drittel aller teilnehmenden Versuchspersonen phänomenale Zustände zu verschaffen, von denen sicher viele Philosophen sagen würden, das sie logisch unmöglich sind: das *Sehen von rötlichem Grün* und das *Sehen von gelblichem Blau*.[95] Mit der Hilfe eines komplizierten Gerätes (einem *eye-tracker*), das die ständigen Mikrobewegungen des menschlichen Auges effektiv neutralisiert, kann man ein stabiles Netzhautbild erzeugen.[96] Präsentiert man nun einer Gruppe von Versuchspersonen als Zielobjekt zwei benachbarte Balken, zum Beispiel einen roten und einen grünen, und stabilisiert die Grenze zwischen den Balken (aber nicht den Rest des Bildes) auf der Netzhaut, dann verschwindet diese Grenze nach einigen Sekunden aus dem Blickfeld der Versuchspersonen. Das visuelle System komplettiert nun automatisch die Farbpräsentation des visuellen Feldes (wie auch gerade jetzt, während sie dies lesen, im Bereich ihres „blinden Flecks“) – aber aufgrund zweier gegensätzlicher Informationen. Bei einem Teil der Versuchspersonen führt dies dazu, daß zwei üblicherweise inkompatible Formate gleichzeitig aktiviert werden: Diese Personen sehen ein rotes und ein grünes Feld, die in der Mitte durch einen Bereich *homogenen, rötlichen*

Hilfe von hohen Primzahlen und besitzen ein Gedächtnis in einer geradezu extremen Größenordnung. *„Und wenn man sie fragt, wie sie so viel in ihrem Gedächtnis bewahren können – eine dreihundertstellige Zahl oder die Milliarde Ereignisse von vier Jahrzehnten – so sagen sie ganz einfach: ‚Wir sehen es.‘“* (Sacks 1987: 260)

[93] Richard Rorty hat wiederholt darauf hingewiesen, daß auch die Zuschreibungskriterien für mentale Zustände geschichtliche Entitäten sind, die nach einer neurowissenschaftlichen Revolution gegen den Cartesianismus zusammen mit dessen theoretischen Entitäten *eliminiert* werden könnten. Er argumentiert für die These, daß alle deskriptiven und explanatorischen Funktionen von Aussagen über psychische Zustände (auch in der Selbstzuschreibung) im Rahmen einer monistischen Theorie erhalten bleiben würden. *„Die Behauptung ‚Niemand hat je einen Schmerz verspürt‘ ist nicht absurder als die Behauptung ‚Niemand hat je einen Dämon gesehen‘, wenn wir eine passende Antwort auf die Frage haben: ‚Was habe ich berichtet, als ich sagte, ich verspürte einen Schmerz?‘ Auf diese Frage könnte die Wissenschaft der Zukunft antworten: ‚Du hast das Auftreten eines bestimmten Gehirnprozesses berichtet, und es würde das Leben für uns einfacher machen, wenn Du in Zukunft sagen würdest: ‚Meine C-Fasern feuern‘, statt zu sagen: ‚Ich habe Schmerzen‘.“* (Rorty 1981a [1965]: 98) Was die subjektive Innerlichkeit der Innenperspektive angeht, die das Thema dieser Arbeit ist, ist diese optimistische Prognose nicht sehr überzeugend: Vielleicht kann man Medizinmänner davon überzeugen, daß es keine Dämonen gibt – aber die Dämonen *selbst*? Vgl. Rorty 1981a, 1981b, 1981c.

[94] Vgl. Dennett 1987: 6.

[95] Vgl. Crane/Piantanida 1983; außerdem das Vorwort von Arthur Danto in Hardin 1988: xf. und Hardin 1988: 124f.

[96] Vgl. Hardin 1988: 125.

Grüns verbunden werden. Auch was das „logisch inkonsistente" dyadische Farbprädikat „gelbliches Blau" angeht, kann man bei einem Teil der Versuchspersonen die entsprechenden phänomenalen Zustände erzeugen.

Von großem Interesse für eine empirisch fundierte Theorie des Geistes sind auch *Synästhesien*. Synästhesien sind polymodale Fusionen qualitativen Gehalts, die als Folge von Schläfenlappenanfällen, Migräne, der Einnahme von psychoaktiven Substanzen, direkter Elektrostimulation des Gehirns und anderer außergewöhnlicher Zustände auftreten können.[97] Es gibt jedoch auch Personen, die von Geburt an Synästheten sind. Solche Personen fühlen Schmerzen in Formen[98], erleben Worte geschmacklich oder in Kombination mit phänomenalen Farben[99], sehen den blauen Geruch von *weißer* Farbe[100], nehmen Geschmackserlebnisse subjektiv als Objekte mit Geruch, Form, Oberflächeneigenschaften und Temperaturen wahr[101], erleben die Note *As* als lavendelfarben, den Akkord As-Dur dagegen als lavendelfarben, violett und weiß mit einem bißchen blau und grün[102] oder den Geschmack von Grüner Minze wie fühlbare, kühle Glassäulen – während Zitronengeschmack sich für manche von ihnen anfühlt, als ob sie die Hand auf ein Nagelbett legen.[103] Am bekanntesten ist das Sehen von Musik:

> When I listen to music I see colored shapes. If I am tired at the end of the day the shapes seem very near. They are always in color. Shiny white isoceles triangles, like long sharp pieces of broken glass. Blue is a sharper color and has lines and angles, green has curves, soft balls, and discs. It is uncomfortable to sit still. I feel the space above my eyes is a big screen where this scene is playing.[104]

Diese Phänomene haben seit etwa 200 Jahren die Aufmerksamkeit der Wissenschaft auf sich gezogen.[105] Heutige Erklärungsversuche gehen von einer polymodalen Kombination von Repräsentaten in relativ frühen Stadien der Informationsverarbeitung im Gehirn aus, während man früher eine Dedifferenzierung des Informationsflusses, die Überkreuzung sensorischer Kanäle oder die Existenz eines synästhetischen Mediators (in Gestalt einer emotionalen oder perzeptuellen Abstraktion, die die spezifischen sen-

---

[97] Im Folgenden stütze ich mich auf Cytowic 1989.
[98] Vgl. Cytowic 1989: 35.
[99] Vgl. Cytowic 1989: 34.
[100] Vgl. Cytowic 1989: 24.
[101] Vgl. Cytowic 1989: 24.
[102] Vgl. Cytowic 1989: 47f.
[103] Vgl. Cytowic 1989: 1.
[104] Die Person ML, zitiert nach Cytowic 1989: 35.
[105] Einen Überblick über die bisher entwickelten Theorien gibt Cytowic 1989, Kapitel 3. Man kann annehmen, daß die hier angesprochene Klasse subjektiver Erlebnisse auch vielen mythologischen, metaphysischen und esoterischen Theorien (wie zum Beispiel in der Theo- oder Anthroposophie) über das „Hellsehen" oder „Aurasehen" zugrundeliegen. Denn tatsächlich sind solche durch neurologische Anomalien hervorgerufenen „übersinnlichen" Erlebnisse dramatische und erlebnismäßig unhintergehbare Bereicherungen der inneren Wirklichkeit: Synästhesien verleihen der jeweiligen Person außergewöhnliche phänomenale Zustände.

sorischen Elemente bis zu einem gewissen Grad ausfiltert) in theoretische Überlegungen mit einbezog. Was kann man aus solchen Phänomenen lernen? Man kann schließen, daß die Formate, in denen unser Gehirn sensorische und emotionale Daten präsentiert, *plastisch* sind und auch ineinander integriert werden können. Man muß außerdem zur Kenntnis nehmen, daß es starke, physiologisch bedingte Idiosynkrasien bezüglich des qualitativen Aspekts subjektiver Zustände geben kann und daß unsere Intuitionen bezüglich des Innenlebens unserer Mitmenschen nur sehr begrenzt zuverlässig sind. Vor allem jedoch muß man *theoretische* Intuitionen bezüglich Qualia in Frage stellen.

In Abschnitt 2.3.2 werde ich einige abweichende phänomenale Welten vorstellen, die unter anderem dem Zweck der weiteren Verunsicherung und Manipulation meiner Leser dienen sollen. Eines der Beispiele wird Blindsicht sein: Das nach gewissen Hirnverletzungen zu beobachtende erfolgreiche Verarbeiten visueller Information ohne die Begleitung entsprechender phänomenaler Zustände. Es gibt nämlich Fälle, in denen Personen in gewissem Ausmaß Farben unterscheiden können ohne dabei ein subjektives Seherlebnis zu haben.[106] Das heißt: Es gibt in manchen Fällen funktional aktive Farbpräsentate – innere Zustände, die einem System diskriminatorisches Verhalten ermöglichen – ohne daß es zu bewußten Farberlebnissen kommt.[107] Wenn das aber so ist, dann müssen wir eine Erklärung dafür finden, was zusätzlich zu mentaler Präsentation geschehen muß, damit sensorische oder emotionale Qualitäten zu Bewußtseinsinhalten werden können. Dies werde ich in dem folgenden Abschnitt versuchen. Lassen Sie uns jedoch vorher noch die zwei wichtigsten Einsichten bezüglich des Phänomens der mentalen Präsentation festhalten. **Erstens** werden die meisten unserer alltagspsychologischen Vorannahmen und die klassischen theoretischen Intuitionen bezüglich Qualia durch neueres empirisches Material *ad absurdum* geführt. **Zweitens** ist mentale Präsentation ein *biologisches* Phänomen: Eine in gewissen natürlich entstandenen Repräsentationssystemen auftretende Form von Informationsverarbeitung.

2.1.4 Subsymbolische Meta-Repräsentation: Bewußtsein

Wir kommen nun zum dritten großen Rätsel für eine nicht-metaphysische Theorie des psychologischen Subjekts. Selbst wenn ich verstanden habe, wie manche meiner inneren Zustände qualitativen Gehalt haben können und wie es dazu kommt, daß sie *meine* sind, bin ich immer noch mit der notorischen Frage nach ihrer *Bewußtheit* konfrontiert. Der anti-naturalistische Standardeinwand gegen jede mechanistisch-funktionalistische Analyse mentaler Phänomene wird häufig eine Variante des folgenden Gedan-

[106] Vgl. Pöppel 1987.
[107] Vgl. hierzu Wilkes 1984: 234f.

kens sein: *„Aber alles, was Du so beschreibst, kann es auch* ***ohne Bewußtsein*** *geben. Wir könnten ein künstliches System so programmieren, daß es die entsprechende innere Konfiguration aufweist und trotzdem gäbe es keinen Grund, anzunehmen, daß dieses System Bewußtsein instantiiert!“* Typischerweise wird mit solchen Einwänden kein genauer Bewußtseinsbegriff mitgeliefert, aus dem hervorginge, worin genau die *a-priori*-Irreduzibilität des fraglichen psychischen Phänomens bestehen könnte.

Seriöse Theoretiker werden sich zumeist auf eine gemeinsame Diagnose als Fundament für eine fruchtbare Diskussion einigen können: *Das Explanandum* ist nicht klar. Wir haben noch nicht verstanden, auf welche Eigenschaft(en) mentaler Repräsentate wir uns beziehen, wenn wir über ihre Bewußtheit sprechen.[108] Das primäre philosophische Problem resultiert aber klarerweise aus der Homogenität[109] phänomenalen Bewußtseins: Wie kann eine Theorie mentaler Repräsentation die einheitliche und bruchlose Natur des subjektiven Erlebnisraums verständlich machen? Wie ist es möglich, daß aus den verschiedenen vom menschlichen Gehirn benutzten Repräsentatformaten ein nicht-fragmentiertes Bewußtsein entsteht, dessen Inhalt unter den Standardbedingungen des Wachzustandes *ein* phänomenales Ich in *einer* Welt ist?

Interessanterweise ist Homogenität auch eine der zentralen phänomenologischen Eigenschaften von Qualia, die diese als primitive und absolut analyseresistente Elemente des Mentalen erscheinen lassen: Jeder Teil eines roten Objekts ist auch rot, jedes Zeitsegment einer inneren Episode vom Typ „süßes Geschmackserlebnis“ zeichnet sich durch dieselbe sprachlich nur sehr schwer kommunizierbare Charakteristik aus. Homogenität ist eines der stärksten Kennzeichen des nicht-relationalen Aspekts von Subjektivität, weil sie das wohl wesentlichste Merkmal des qualitativen Gehalts mentaler Zustände ins Blickfeld rückt. Somit wird Homogenität bereits bei der Betrachtung einzelner „Bausteine“ der phänomenalen Wirklichkeit zum Problem – und zwar, weil unklar ist, wie diese subjektive Einheitlichkeit[110] das Resultat von Myriaden von Informationsverarbeitungsereignissen auf der neuronalen Ebene sein könnte. Notorische Optimisten des

---

[108] Gute Einführungen in die Problematik des Bewußtseinsbegriffs geben Allport 1988 und Bieri 1991. Die These, daß das Prädikat „bewußt“ in einer wissenschaftlichen Taxonomie mentaler Zustände nur eine sehr untergeordnete Rolle zu spielen hat, vertritt Wilkes 1988. Vgl. auch Metzinger 1995 a, b, c.

[109] Eine interessante Diskussion des Sellarsschen *Grain-Arguments* bietet Lycan 1987: 93ff. Vgl. auch Metzinger 1995 c.

[110] Die subjektive Einheitlichkeit als Qualität einzelner mentaler Inhalte oder des inneren Erlebnisraums als Ganzem führt in cartesianischer Interpretation zur Einheit des Subjekts, also von der Eigenschaftsanalyse zur Objektbildung. Homogenität ist aber nicht dasselbe wie Descartes' Unteilbarkeit des Ich. Wie ich weiter oben bereits gesagt habe, sind egologische und substantialistische Theorien des Mentalen in Schwierigkeiten geraten durch empirische Resultate von Untersuchungen an kommissurotomierten Patienten. Es scheint jedoch plausibel, daß zumindest die den Aktivitäten der dominanten Hemisphäre postoperativer Patienten korrespondierenden mentalen Repräsentate (denn nur über sie liegen uns in der Experimen-

kybernetischen Paradigmas warten an dieser Stelle häufig mit dem Zauberwort „Komplexität“ auf und vermuten, daß ab einer bestimmten Komplexitätsstufe gewisse emergente Qualitäten plötzlich und nicht prognostizierbar einfach *auftreten*. In Abwesenheit eines klar ausgearbeiteten und auf den psychophysischen Problembereich zugeschnittenen Emergenzbegriffs[111] führt diese Form obskurantistischer Zuversicht jedoch nicht weiter. (Die skeptisch-konservative Variante ist der *pessimistische* Obskurantismus: Er postuliert für jede problematische Eigenschaft mentaler Zustände starke nicht-physikalische Entitäten und behauptet in Ermangelung einer präzisen Beschreibung dieser Entitäten vorerst einmal flugs ihre Irreduzibilität.)

Betrachtet man wiederum unser phänomenales Feld als Ganzes, dann werden wir ein zweites Mal und in verschärfter Form mit dem oben angedeuteten Aspekt des Qualia-Problems konfrontiert. Auch unser Bewußtsein *als Ganzes* zeichnet sich durch eine eigentümliche Homogenität aus. Zwar treten innerhalb unseres Bewußtseins Inhalte mit sehr verschiedenem qualitativen Gehalt auf, aber es gibt eine durchgängige Qualität, durch die alle mentalen Repräsentate – seien es Repräsentate von Objekten aus der Systemumwelt oder Repräsentate von Systemeigenschaften *als* mental – miteinander verknüpft werden. Diese homogene Eigenschaft, die allen Bausteinen unseres repräsentationalen Gesamtzustandes ebenso zukommt wie allen Teilen eines roten „Bewußtseinsgegenstands“ die phänomenale Röte und die das Leib-Seele-Problem zu einem so vertrackten und schwer zu behandelnden Problem macht, stellt vielleicht das Kardinalproblem der Philosophie des Geistes dar: *Bewußtheit*.

Bewußtheit ist diejenige Eigenschaft, die eine Teilmenge mentaler Repräsentate auszeichnet und allen Elementen dieser Menge eine prägnante, introspektiv unbezweifelbare homogene Gesamtqualität verleiht. Bewußtheit schwankt in ihrer phänomenalen Intensität und die notwendigen biologischen Randbedingungen für ihre Entstehung werden immer deutlicher.[112]

---

talsituation verbale Introspektionsberichte vor) wieder durch homogene Bewußtheit charakterisiert sind. Auch wenn ein funktionales, selbstrepräsentierendes System gespalten wird, bleibt also die Homogenität der ihm nun noch möglichen phänomenalen Zustände bestehen.

[111] An einem unklaren Begriff von „Emergenz“ scheitern eine ganze Reihe zeitgenössischer Leib-Seele-Theorien, etwa der emergentistische Materialismus von Bunge 1980 (1984), der in Wirklichkeit ein biologistischer Reduktionismus ist (vgl. Metzinger 1985: 237ff), die interaktionistische Drei-Welten-Lehre von Popper/Eccles 1982 (vgl. Metzinger 1985: 261ff) oder die „symmetrische Typ-Emergenz“ von Hastedt 1988, die eher an einen dualistischen Interaktionismus mit Primat des Physikalischen erinnert (vgl. Metzinger 1990). Einen guten Überblick über die Perspektiven eines nicht-reduktiven Materialismus und den gegenwärtigen Diskussionsstand gibt der von Beckermann/Flohr/Kim 1992 herausgegebene Sammelband. Vgl. auch Stephan 1999.

[112] Die Homogenität von Bewußtheit scheint andererseits ein „digitales“ Phänomen zu sein: Wenn Bewußtheit überhaupt entsteht, dann ist sie auch homogen über die jeweilige Menge von Repräsentaten verteilt. Dagegen ist die jeweilige Auswahl der bewußten mentalen Repräsentate selbst wiederum hochgradig variabel.

Aus empirischer Perspektive[113] mag sie als eine teilweise starken Schwankungen unterworfene, umfassende Hintergrundvariable erscheinen. In theoretischer Hinsicht taucht sie auf allen Ebenen der Analyse erneut in Form einer modalen Intuition auf, die Peter Bieri einmal als die „tibetanische Gebetsmühlenfrage" bezeichnet hat. Welches mentale Phänomen auch immer einer physikalistischen oder funktionalistischen Analyse unterzogen wird, am Ende aller theoretischen Bemühungen taucht wieder die unvermeidliche, oben bereits angesprochene Frage auf: *„Aber wäre all dies nicht auch ohne Bewußtsein möglich?"*

Es hat in der neueren Diskussion viele Versuche gegeben, das mysteriöse Phänomen einzukreisen. Dabei ist auf vielen Beschreibungsebenen – von den physikalischen Eigenschaften des Gehirns bis hinauf zu seiner repräsentationalen Aktivität – nach hinreichenden und notwendigen Bedingungen für die Zuschreibung des Prädikats „bewußt" gesucht worden. Lassen Sie uns zunächst einen kurzen Blick auf drei Versuche von Hirnforschern werfen, die rätselhafte „Feldqualität" auf der neurobiologischen Beschreibungsebene dingfest zu machen. Das Ziel einer harten operationalistischen Strategie ist es zum Beispiel gewesen, den Begriff des „bewußten Zustands" relativ zu genau spezifizierten Arten von Experimenten zu definieren (durch den schönen Ausdruck $\alpha\,(D_{Ep}^{e+})^6$), indem man verlangt, daß die Versuchsperson sich *„einer Eigenschaft gewahr ist und daß sie sich gewahr sein von sich nicht gewahr sein mentalistisch unterscheiden kann."*[114] Den Versuchen von Neurowissenschaftlern, aufgrund empirischer Erkenntnisse zu einer genaueren Begriffsbestimmung zu kommen, ist gemeinsam, daß sie eine starke pragmatische Komponente aufweisen. Ernst Pöppel hat in Zusammenhang mit Untersuchungen zur zeitlichen Quantelung der Infomationsverarbeitung im Gehirn eine Definition des Zustands „bewußt" vorgeschlagen, die sich aus neuropsychologischen Erkenntnissen bezüglich des „subjektiven Gegenwartsfensters" ergeben. Dieses Gegenwartsfenster wird vom menschlichen Gehirn geöffnet, indem es eine Vielzahl von Operationen zu maximal drei Sekunden andauernden phänomenalen Wahrnehmungsgestalten integriert. Pöppel sagt:

> Faßt man die Befunde über die Klassifikation psychischer Phänomene und die Beobachtungen über zeitliche Integration zusammen, bietet sich eine pragmatische Definition des Zustands „bewußt" an; es läßt sich sagen: "Bewußt" ist jener Zustand, bei dem für jeweils wenige Sekunden aufgrund eines integrativen Mechanismus des Gehirns „Mentales" repräsentiert ist, d. h. im Fokus der Aufmerksamkeit steht.[115]

[113] Einen ausgezeichneten Überblick über den Status von „Bewußtsein" in der gegenwärtigen philosophisch-wissenschaftlichen Theorienbildung gibt der Sammelband von Marcel/Bisiach 1988. Vgl. auch Globus/Maxwell 1976. Siehe auch Metzinger 1995 a.

[114] Dies ist nur der Grundgedanke. Die genaue Definition findet sich in Werth 1983: 119.

[115] Vgl. Pöppel 1989: 30; dazu auch Pöppel 1985.

In der empirischen Forschung haben solche Definitionen ihre Berechtigung.[116] Philosophen werden mit ihnen wenig glücklich sein, weil mit Ausdrücken wie „mentalistisch" oder „Mentales" und besonders mit „Aufmerksamkeit" (als *gerichteter Bewußtheit*) der zu definierende Begriff als Teil des Definiens wieder mit eingeschmuggelt wird. Trotzdem hat gerade die zuletzt genannte Definition mit der zeitlichen Homogenität subjektiver Erlebnisse ein zentrales Merkmal des Phänomens „Bewußtsein" im Auge. Die empirischen Entstehungsbedingungen von Eigenschaftsbindung und Metarepräsentation nimmt eine andere interessante Hypothese ins Visier, deren Ziel die Fixierung der notwendigen und hinreichenden neurobiologischen Bedingungen für phänomenales Bewußtsein beim Menschen ist. Sie stammt von Hans Flohr und basiert auf neueren Erkenntnissen über selbstorganisierende assoziative Netzwerke. Die Entdeckung eines molekularen Mechanismus (des NMDA-Rezeptor-Komplexes[117]), der als Grundlage eines besonders *schnellen* Typs von Synapsen – den schon länger vermuteten „schnellen Hebb-Synapsen" oder „Malsburg-Synapsen" – fungieren könnte, hat es möglich gemacht, von einer spontanen und sehr schnellen Bildung von Zellverbänden (*„neural cell assemblies"*[118]) auszugehen, deren Formierung ohne permanente Veränderung synaptischer Gewichtungsfaktoren vonstatten gehen kann. Die technischen Details interessieren hier nicht weiter. Aus philosophischer Perspektive wird allerdings der sich aus ihnen ergebende theoretische Kandidat für die Zuschreibung von phänomenalem Bewußtsein die Aufmerksamkeit erwecken, nämlich die *Rate der pro Zeiteinheit gebildeten Zellverbände*. Hier die Flohrsche Hypothese:

> The rate at which different types of assemblies are generated determines:
>
> (1.) The amount of representations generated in a given period of time;
>
> (2.) the maximal size of assemblies that can be generated in a given time period and, hence, the maximal complexity of representations. The degree of activation of a neural net therefore determines qualitative differences in what the brain can and cannot represent;
>
> (3.) the duration of representations;
>
> (4.) the amount, complexity and duration of metarepresentations referring to internal states that can be formed per time, and
>
> (5.) the amount and the quality of computational processes possible in a given period of time – for these depend on the amount and the quality of given representations.

---

[116] Weitere Versuche, empirisch adäquate Definitionen von „Bewußtsein" herauszukristallisieren finden sich in Bisiach 1988: 102f, Marcel 1988: 128–131. Weiskrantz 1988: 183f weist darauf hin, daß die Bedeutung strikter Definitionen in der Forschungspraxis nicht überschätzt werden sollte. Eine verwandte Argumentation für die Lokalisierung provisorischer Definitionen in empirisch motivierten Theorien und gegen philosophische Apriorismen findet sich bei Frau Churchland. Vgl. Churchland 1988: 281ff.

[117] Genauere Angaben über die empirischen Daten und weitere Literaturangaben finden sich in Flohr 1991 und Kennedy 1989.

[118] Vgl. Hebb 1949.

> My hypothesis is this: deficits in phenomenal consciousness as well as certain functional deficits occur if the rate of assembly formation falls below a critical threshold level. Whenever this level is surpassed, phenomenal states must necessarily occur. A critical production rate of neural assemblies is the necessary and sufficient condition for the existence of phenomenal states. Awareness is the result of the system's capacity to actively generate representations and metarepresentations.[119]

Die Flohrsche Hypothese ist ein gutes Beispiel für Versuche von Neurowissenschaftlern, das psychische Globalphänomen „Bewußtsein" einzukreisen und seine notwendigen und hinreichenden biologischen Bedingungen immer genauer zu formulieren. Von Interesse für eine Theorie des Menschen als eines bewußten Wesens sind auch neuere Überlegungen zu der Hypothese, daß eine Phasengleichheit neuronaler Oszillationen das neurobiologisch-funktionale Substrat dessen sein könnte, was wir auf der repräsentationalen Ebene als *Eigenschaftsbindung* bezeichnen und auf der phänomenalen Ebene als den Holismus unseres Bewußtseinsraums und einzelner seiner Elemente erleben.[120] Wenn solche Projekte einmal erfolgreich sein sollten, werden wir über empirisch fundierte Zuschreibungskriterien für das Prädikat „bewußt" beim Menschen verfügen. Dagegen scheint das philosophische Projekt einer *generellen* Theorie des Geistes zu verlangen, daß wir „Bewußtsein" als eine mentale *Universalie* rekonstruieren:[121] Was ist es, das Bewußtsein bei allen möglichen Systemen, die die entsprechenden Zustände instantiieren können, zu *Bewußtsein* macht? Denn als Philosophen wollen wir ja gerade wissen, wie eine möglichst umfassende, d. h. spezies- und *hardware*-unabhängige Theorie der Subjektivität aussehen müßte.[122]

Viele Funktionalisten haben deshalb versucht, abstraktere Eigenschaften informationsverarbeitender Systeme zur Definition von „Bewußtsein" einzusetzen. Ein solches Unternehmen kommt dem alten philosophischen Ziel einer Hardware- und speziesunabhängigen universellen Psychologie näher, läuft aber ständig Gefahr, zu liberal bei der Entwicklung von Zuschrei-

---

[119] Vgl. Flohr 1991: 255; dazu auch Flohr 1992: 54.

[120] Einen interessanten Versuch in dieser Richtung stellt zum Beispiel die Prototheorie von Francis Crick und Christof Koch dar, vgl. Crick/Koch 1990 a,b sowie Crick 1984. Diese Ansätze gehen über die quantitative Hypothese von Flohr hinaus, die ich hier als Beispiel gewählt habe, weil sie einen spezifischen Mechanismus der Eigenschaftsbindung durch temporale Kodierung im Auge haben. Vgl. dazu etwa Barinaga 1990, Engel et al. 1991a, b, c, 1992a, b, c, Gray et al. 1989, Kreiter/Singer 1992, Singer 1989, 1990, 1993, (im Druck) und Fußnote 159. Siehe auch Metzinger 1995 b.

[121] Es könnte eben dieses Projekt sein, daß im Verlauf einer Naturalisierung des psychologischen Subjektbegriffs letztlich aufgegeben werden muß – weil der Begriff des Bewußtseins im Verlauf seiner semantischen Anreicherung durch die empirischen Wissenschaften auf immer „tiefere" Beschreibungsebenen abwandert und dadurch seine Universalität verliert.

[122] Eine naturalistische Theorie des Subjekts muß deshalb keine biologische Theorie sein: Wenn Subjektivität ein repräsentationales Phänomen ist, dann können auch nicht-biologische (z. B. postbiotische) Systeme unter Ausnutzung physikalischer Gesetzmäßigkeiten psychologische Subjekte erzeugen. Vgl. etwa Langton 1989.

bungskriterien für „Bewußtsein“ zu sein.[123] Hier ein typischer Versuch, „Bewußtheit“ durch ein auf einer höheren, nicht-physikalischen Beschreibungsebene angesiedeltes Abstraktum – nämlich als Eigenschaft einer bestimmten Klasse von Parallel-Algorithmen – auszuzeichnen:

> Consciousness may originally have emerged as a processor promoted to be an operating system and to monitor other processors so as to override pathological configurations. The division between the conscious and unconscious components of the mind bears out this distinction: the contents of consciousness appear to be the values of the parameters that serve as input and output to the operating system. The particular subjective experience of consciousness depends on self-awareness, which arises from both the recursive ability to embed mental models within mental models and the mind's possession of a high-level model of the capabilities of his own operating system.[124]

Die Bewußtseinstheorie von Johnson-Laird geht von einer Vielzahl parallel arbeitender Prozessoren aus, deren Aktivität unbewußt ist. Ein übergeordnetes Betriebssystem hat die Aufgabe, den Output dieser Prozessoren so zu organisieren, daß eine geordnete Abfolge von Verhaltensweisen (ein *serielles* Muster) erzeugt werden kann.[125] Außerdem ist es die Aufgabe des Betriebssystems, „festgefahrene“ Prozessoren voneinander zu lösen. Die Bewußtseinsinhalte sind die gegenwärtigen Werte der Parameter des Betriebssystems: Bewußtsein ist ein Repräsentationsphänomen, das innerhalb eines Parallelsystems durch die Notwendigkeit einer Überwachung, Kontrolle und Kanalisierung des Outputs erzeugt wird.[126] Repräsentationale und kognitivistische Bewußtseinstheorien verschieben auf diese Weise häufig das Leib-Seele-Problem eine Ebene nach oben und werfen dann die Frage nach

---

[123] Dieses *Liberalismus-Chauvinismus-Dilemma* ist ein grundsätzliches Problem des Funktionalismus in der Philosophie des Geistes. Ned Block hat diesen Punkt herausgearbeitet; vgl. Block 1978. Die Entwicklung eines Teleofunktionalismus ist eine der gegenwärtig interessantesten Strategien, dem Problem zu begegnen. Vgl. etwa Lycan 1987, Sober 1985; einen guten Überblick gibt Lycan 1990: 59ff.

[124] Vgl. Johnson-Laird 1983: 477; die Ziele und Möglichkeiten einer computationalen Theorie des Bewußtseins erläutert Johnson-Laird in Johnson-Laird 1988.

[125] Den Gedanken, daß Bewußtsein durch die Emulation eines seriellen Prozessors *auf einem PDP-System* entstanden sein könnte, hat häufig Dennett vertreten: *„Human consciousness is itself a huge complex of memes (or more exactly, meme-effects in brains) that can best be understood as the operation of a „von Neumanesque“ virtual machine implemented in the parallel architecture of a brain that was not designed for any such activities. The powers of this virtual machine vastly enhance the underlying powers of the organic hardware on which it runs,. . .“* (Dennett 1991: 210). Die Erzeugung „sequentieller Formate“ und die Lösung von Symbolverarbeitungsaufgaben auf einem Parallelrechner kann prinzipiell durch verschiedene Strategien verwirklicht werden, etwa durch Hybrid-Systeme, z. B. ein Netz oder Mengen von Netzen, die mit einem symbolverarbeitenden Controller verbunden sind; vgl. McClelland/Rumelhart 1986: 548f.

[126] Die Beziehung zwischen phänomenalem Bewußtsein und Metarepräsentationsfunktionen sowie die Schwierigkeiten bei der Zuschreibung kausaler Eigenschaften in bezug auf diese Entitäten in den theoretischen Modellen von Shallice und Johnson-Laird untersucht Bisiach 1988: 114f.

der Beziehung zwischen phänomenalem Bewußtsein und dem *Computational Mind* auf.

Es hat in letzter Zeit auch Versuche gegeben, Bewußtheit als eine Form von mentalem Gehalt höherer Ordnung zu analysieren, und dabei jeden Rekurs auf die Terminologie des Informationsverarbeitungs-Paradigmas zu vermeiden. Damit macht man (ähnlich wie ich zu Beginn dieses Kapitels vorgeschlagen habe) Bewußtsein zu einer relationalen Eigenschaft mentaler Zustände, die ihnen nicht wesensmäßig zu eigen ist (wie nach dem intuitiven Denkmodell „selbstleuchtender innerer Episoden"), sondern ihnen durch einen sie begleitenden *Gedanken höherer Ordnung*[127] verliehen wird. Man kann deshalb auch von *transitivem* Bewußtsein sprechen.

> We are conscious of something, on this model, when we have a thought about it. So a mental state will be conscious if it is accompanied by a thought about that state. The occurence of such a higher-order thought makes us conscious of the mental state; so that state is a conscious state. Similarly, when no such higher-order thought occurs, we are unaware of being in the mental state in question, and the state is then not a conscious state. The core of the theory, then, is that mental states are conscious states when, and only when, they are accompanied by a suitable higher-order thought (HOT).[128]

Dieser hier im Vokabular der *folk psychology* formulierte Grundgedanke taucht in vielen kognitionswissenschaftlichen und philosophischen Entwürfen der Gegenwart in verschiedenen Varianten auf: Bewußtsein entsteht dadurch, daß bereits repräsentationalen Gehalt tragende innere Zustände ihrerseits zum Gehalt höherstufiger, *metarepräsentationaler* Zustände werden. Ray Jackendoff hat eine solche differenzierte *Representational Theory of Mind* entwickelt, die auch eine Reihe empirischer und klinischer Phänomene erklärt. Für ihn ist Bewußtsein ein Epiphänomen, eine Projektion funktionaler Zustände ohne kausale Wirksamkeit. Er geht folgerichtig in seiner *Intermediate-Level Theory* davon aus, daß die Inhalte phänomenalen Bewußtseins einer *mittleren* Ebene der Repräsentationshierarchie entsprechen – etwa dem Kurzzeitgedächtnis, aus dem sie ausgewählt und durch Meta-Repräsentation verstärkt werden.

> The distinctions of form present in each modality of awareness are caused by/supported by/projected from a structure of intermediate level for that modality that is part of the matched set of short-term memory representations designated by the selection function and enriched by attentional processing.[129]

Das Problem funktionalistischer und neuroinformatischer Definitionen von „Bewußtsein" liegt darin, daß an den unter diesen Begriff fallenden mentalen Operationen jeweils eine sehr grosse Anzahl von Subsystemen

[127] Problematisch an solchen höherstufigen Gedanken ist meiner Auffassung nach ihr *assertorisches* und *propositionales* Moment. Ihr Gehalt muß sein, daß *man selbst* sich *jetzt* in diesem speziellen mentalen Zustand befindet. Vgl. dazu Rosenthal 1990: 27.

[128] Vgl. Rosenthal 1990: 27; dazu auch Dennett 1991: 307ff.

[129] Vgl. Jackendoff 1987: 298.

beteiligt ist: Welche Funktion ist die *kritische* Funktion – der Input der Sprachzentren?[130] Sind Bewußtseinsinhalte Datenstrukturen im Puffer des Kurzzeitgedächtnisses[131], die Prozesse bei der Selektion mentaler Schemata und Zielvariablen[132] oder Aktivitäten eines Überwachungssystems?[133] Das Problem besteht immer darin, unscharfe mentale Universalien der Alltagsphänomenologie (bzw. klassischer, spekulativ-idealistischer Theorien des Geistes) auf sinnvolle Weise[134] mit dem erst im Entstehen befindlichen Begriffsapparat einer am Modell der Informationsverarbeitung orientierten theoretischen Psychologie zu verknüpfen.

> First, the conceptual systems of phenomenology and information-processing/ computational psychology are sufficiently loose and remote from each other that one may be able to produce apparently plausible correspondences to ‚consciousness' from a variety of loci within information-processing theories. One should therefore be suspicious of mechanistic correspondences for ‚consciousnes' which are based on broad information-processing theories and a limited range of phenomenological generalizations
> The other general conclusion is that it may not be possible to produce a satisfactory correspondence for all the phenomenological generalizations that can be advanced with the operation of a single component within an information-processing model. Yet this would appear to be the explicit or implicit aim for most previous attempts to explain ‚consciousness' from an information processing or functionalist approach.[135]

Generell haben mechanistische Ansätze zu einer Theorie des Geistes das Problem, phänomenale Ganzheiten – wie sie durch subjektive Erlebnisqualitäten und Bewußtheit (Bewußtheit ist die umfassendste dieser Ganzheiten) dargestellt werden – auf Elemente tieferliegender naturwissenschaftlicher Beschreibungsebenen und deren Relationen zu reduzieren oder anderweitig[136] in Beziehung zueinander zu setzen. Denn in der Tat scheint es unverständlich, wie aus vielen Milliarden von funktional miteinander vernetzten Mikroereignissen im Bereich biologischer Informationsverarbeitung atomische, homogene und introspizierbare Makro-Eigenschaften entstehen sollen. Man muß jedoch bei solchen Fragestellungen sehr vor-

---

130 Vgl. Dennett 1969.

131 Einen Überblick über solche am Begriff der Informationsverarbeitung orientierten Modelle gibt Tim Shallice, auf ihn stütze ich mich hier. Vgl. Shallice 1988.

132 Vgl. Shallice 1972, 1988: 320.

133 Vgl. Mackay 1966, Marshall/Morton 1978, Mandler 1985.

134 Das heißt auf eine Weise, die heuristisch fruchtbar und offen für ständige empirische Anreicherungen ihres semantischen Gehalts ist, die also befriedigende metatheoretische Perspektiven auf die bekannten philosophischen Probleme eröffnet und dabei gleichzeitig neue testbare Hypothesen erzeugt.

135 Vgl. Shallice 1988: 325.

136 Nicht-reduktive Relationstypen wären etwa *Supervenienz* oder *Emergenz*. Sie sind bereits mit sehr unterschiedlichem Erfolg auf das Leib-Seele-Problem angewandt worden. Vgl. Bunge 1984, Haugeland 1982, Kim 1974, 1978, 1982, Stoecker 1992: Kapitel 7 und 8.

sichtig sein. Die Mikro-Ereignisse, von denen wir nicht verstehen, wie sie Grundlage phänomenalen Bewußtseins sein können und über die wir mehr und mehr erfahren, werden ausschließlich aus der („Dritte-Person"-) Perspektive der Neurowissenschaftler untersucht und erklärt. Sie gehören weder zu unserer mesokosmischen Lebenswelt, noch sind sie introspizierbar.[137] Was wir ihnen manchmal vorschnell nach der Mikro-Makro-Analogie[138] als subjektive Erlebnisinhalte zuordnen, sind nicht Gegenstände derselben Erkenntnismethode, sondern Resultate der Anwendung des bis dato wissenschaftlich noch unverstandenen Prozesses der mentalen Repräsentation *auf sich selbst*. Ob Introspektion qua mentale Metarepräsentation überhaupt sinnvollerweise als ein Erkenntnisinstrument gelten kann, das zur Konkurrenz oder zum Vergleich mit dem Erkenntnisinstrument „Wissenschaft" in der Lage ist, scheint zudem äußerst fraglich.[139]

Dieser kurze Gang durch den Garten gegenwärtiger Bewußtseinstheorien zeigt, daß man das Explanandum („phänomenales Bewußtsein") auf vielen

[137] Vgl. jedoch Churchland 1985b.

[138] Die Mikro-Makro-Intuition ist auch eines der entscheidenden philosophischen Motive, die zur Entwicklung der Supervenienz–These geführt haben. Wenn wir physikalische Objekte mittlerer Größenordnung betrachten, nehmen wir an ihnen makroskopische Eigenschaften wir Solidität, Farbe oder Form wahr. Diese Makro-Eigenschaften sind nicht durch universell gültige nomologische Korrelationen mit den ihnen zugrundeliegenden mikroskopischen Eigenschaften verknüpft. Da es keine bikonditionalen Verknüpfungen in Gestalt von „Brückengesetzen" universeller Gültigkeit gibt (die Festigkeit und Farbe eines Tisches können aus ganz verschiedenen Materialien resultieren), kann man Makro-Eigenschaften nicht auf Mikro-Eigenschaften reduzieren oder in Begriffen der tieferliegenden Beschreibungsebene definieren. Makro-Eigenschaften erfordern eine eigene Beschreibungsebene und diese ist *nomologisch inkommensurabel* mit der „unter" ihr liegenden molekularen oder atomaren oder quantenmechanischen Beschreibungsebene. Trotzdem glaubt niemand, daß Tische in irgendeinem metaphysisch-ontologischen Sinne eigenständige Entitäten einer anderen Ebene sind. Und schließlich glauben wir, daß die Makro-Eigenschaften von Tischen durch ihre Mikro-Eigenschaften *determiniert* sind – obwohl kausale Erklärungen (die strikte Gesetze instantiieren, vgl. Davidson 1981) nun nicht mehr möglich sind. Viele Philosophen haben deshalb geglaubt, daß man eine Art *minimalen Materialismus* formulieren kann, der die eben geschilderte Intuition auf den Bereich psychophysischer Korrelationen überträgt. Subjektivität ist jedoch mehr als ein von unten determiniertes Makro-Phänomen: Subjektivität als das Auftreten perspektivisch organisierter Repräsentationsräume in Biosystemen bringt eine neue Qualität mit sich, die alle Beispiele, aus denen sich die Mikro-Makro-Intuition speist (Tische, Wasserwellen, Regenbögen), nicht besitzen – nämlich *Innerlichkeit*. – Bezüglich des mit Blick auf diese Problematik und psychologische Eigenschaften entwickelten Erklärungstyps der Instantiierungs-Erklärung vgl. Cummins 1983, Eimer 1990. Eine gute Erläuterung des Begriffs „Supervenienz" in deutscher Sprache bietet Stoecker 1992.

[139] Wie bereits angedeutet liegt dies daran, daß externe, öffentliche Repräsentate in propositionalem Format (wissenschaftliche Theorien) einen völlig anderen Typus von Wissen instantiieren als mentale Meta-Repräsentate (Inhalte von Introspektion; vgl. hierzu Abschnitt 2.2.1). In jedem Falle haben die mechanistische Theorien des Mentalen erzeugenden *scientific communities* eines mit den biologischen Gehirnen der einzelnen Wissenschaftler gemein: Sie sind auf Teile der Welt gerichtete Repräsentationssysteme. Daß die von ihnen generierten Repräsentate nur dann miteinander in Kontakt gebracht werden können, wenn es sich um *propositionale* handelt, darf uns nicht verwundern. Vgl. auch Churchland 1989, Teil II.

Ebenen der Systembeschreibung versuchsweise fixieren kann, wobei das prinzipielle Problem einer Zuordnung der dann entstehenden Termini zur introspektiven Erfahrung und ihrer Interpretation durch das cartesianisch gefärbte Alltagsidiom bestehen bleibt. Welche vorläufige Antwort auf die Bewußtseinsfrage könnte man nun aus der Perspektive einer repräsentationalistischen Theorie des Geistes geben? Zunächst kann man – in Abwandlung des klassischen „Reflexionsmodells" der Subjektivität – die These aufstellen, daß bewußt genau all jene mentalen Simulate, Repräsentate und Präsentate sind, die durch einen zweiten internen Repräsentationsprozess erfaßt und dadurch zu Inhalten phänomenalen Bewußtseins *gemacht* werden.

Mentale Meta-Repräsentation: $\mathbf{M}_{\mathbf{M\text{-}Rep}}$ $(S,X,Y_{1\text{-}n})$

- X repräsentiert $Y_{1\text{-}n}$ *für* S.
- $Y_{1\text{-}n}$ ist eine Teilmenge der in S aktiven mentalen Repräsentate, Simulate oder Präsentate.
- X ist das jeweilige *bewußte* Modell des Selbst und der Welt in und *für* S.
- $M_{M\text{-}Rep}$ ist antisymmetrisch.
- $M_{M\text{-}Rep}$ immer ein Fall von $\mathbf{M}_{\mathbf{Rep}}$.

Wir haben damit eine weitere Differenzierung des ursprünglichen Begriffspaars („mentale Repräsentation" / „mentale Simulation") entwickelt, um eine problematische Eigenschaft phänomenaler Zustände genauer zu analysieren. Den Grundgedanken dieser Begriffsbildung kann man in Form der folgenden provisorischen Hypothese über die Entstehung bewußter Repräsentate in einem informationsverarbeitenden System darstellen.

(MRT): Die Inhalte phänomenalen Bewußtseins sind **Meta-Repräsentate**, die für ein System eine Teilmenge der gegenwärtig in ihm aktivierten mentalen Simulate und Präsentate abbilden. Das, was alle bewußten inneren Zuständen miteinander gemein haben, ist die Tatsache, daß sie durch eine einheitliche *Metarepräsentationsfunktion* erfaßt werden.

Beim gegenwärtigen Stand unseres empirischen Wissens wären Vermutungen über die neurobiologische Realisierung der entsprechenden Metarepräsentationsfunktion im menschlichen Gehirn, die über die eben angesprochenen Prototheorien der Neurowissenschaftler hinausgehen, verfrüht und hochspekulativ. Man kann aber vor dem Hintergrund neuerer Erkenntnisse über massiv parallel arbeitende Systeme und die Art der von ihnen erzeugten internen Repräsentate bereits auf einen Punkt hinweisen, der eine moderne, naturalistische Theorie des Bewußtseins einmal von traditionellen „Reflexionsmodellen" unterscheiden könnte: Eine repräsentationalistische Theorie des Bewußtseins ist nicht unbedingt auch eine *kognitive* Theo-

rie des Bewußtseins.[140] Wenn nämlich die metarepräsentierende Funktion auf dem *subsymbolischem* Niveau eines Systems operiert – etwa, indem die Aktivierungsvektoren einiger innerer Schichten eines neuronalen Netzes durch die Aktivierungsvektoren anderer Schichten abgebildet werden – dann wäre die diesen Prozess beschreibende Theorie des Bewußtseins eine mikrokognitive oder mikrofunktionalistische.[141] Eine solche Theorie des Bewußtseins hätte sich somit an der Feinstruktur eines Netzwerkes zu orientieren, sie würde Bewußtsein als *nicht-regelgeleitete Darstellung einer Menge von Subsymbolen durch eine andere Menge von Subsymbolen* innerhalb eines Parallelsystems erklären. Da das Gehirn ein massiv parallel arbeitendes selbstorganisierendes System ist, liegen solche Vermutungen nahe. Bewußtsein muß nicht auf deklarativem Meta-Wissen eines Systems über seine eigenen inneren Zustände oder auf assertorischen *„Higher-Order-Thoughts"* à la Rosenthal beruhen. Wenn Bewußtsein überhaupt als eine höherstufige, rekursive Form von biologischer Informationsverarbeitung erklärt werden kann, dann ließen sich bereits jetzt aus der Theorie konnektionistischer Systeme theoretische Modelle entwickeln, die zeigen, wie Metarepräsentation ohne diskrete interne Symbole oder propositionales inneres Wissen realisiert werden könnte. Die neue psychologische Eigenschaft der „Bewußtheit" würde in solchen Systemen nicht die Einheitlichkeit des internen Darstellungsraums „von oben" stiften. Vielmehr würde diese umgekehrt einfach dadurch entstehen, daß ein komplexes, nicht-lineares physikalisches System sich bei jedem gegebenen Input wieder in seinen energieärmsten Zustand zu relaxieren versucht und dabei durch Eigenschaftsbindung eine Objektbildung auf höchster Stufe vornimmt – das heißt: ein einheitliches *Weltmodell* erzeugt. Damit gibt es – das ist jenseits aller technischen Details die philosophisch interessante Einsicht – die Möglichkeit eines Mittelwegs zwischen platten Reduktionismen, diffusen Emergenztheorien des Bewußtseins und klassisch-kognitivistischen bzw. sententialistischen Erklärungsstrategien. Dieser Mittelweg heißt subsymbolische Meta-Repräsentation.

Diese Erklärungsstrategie ist das Mittel der Wahl für jede repräsentionalistische Rekonstruktion von Bewußtsein als einer inneren Erfahrung – und als Denkfigur ist sie natürlich keineswegs neu.[142] Darum kann man sich

---

[140] Vgl. Baars 1988.

[141] Vgl. Clark 1987: 34. Eine mikrofunktionalistische Theorie könnte in solchen Fällen, in denen ein System *überhaupt* keinen globalen Output mehr erzeugt, die einzige Möglichkeit sein, ihm noch sinnvollerweise eine psychische Eigenschaft wie „Bewußtheit" zuzuschreiben. Beim Menschen ist eine solche Situation (in der behaviorale Zuschreibungskriterien keine Anwendung mehr finden können) zum Beispiel beim „Totalen *Locked–In*–Syndrom" (TLIS) gegeben, das nach schweren Läsionen im Bereich der Pons und/oder des Mesenzephalons auftreten kann. Zur Substitution behavioraler durch neuropsychologische Zuschreibungskriterien in solchen Situationen vgl. Kurthen 1990: 70ff.

[142] Bezüglich der in den verschiedensten Varianten auch heute immer wieder auftauchenden Akt-Objekt-Äquivokation von „Repräsentation" bzw. „Idee" gab es bereits vor der klassischen deutschen Subjektphilosophie eine philosophische Kontroverse zwischen Arnauld und Malebranche (vgl. Scheerer 1990b: 19ff), in deren Verlauf Arnauld interessanterweise

fragen, ob sie Schwierigkeiten ihrer philosophiegeschichtlichen Vorfahren geerbt hat. Die Rückwendung des Subjekts in sich selbst spielte eine große Rolle in den klassischen Selbstbewußtseinstheorien des spekulativen Idealismus, zum Beispiel in der Fichteschen Egologie. Bei Fichte setzt sich das Ich ohne jede Vermittlung selbst, jedoch bereits in seiner Subjekt-Objekt-Struktur. Diese unmittelbare Selbstsetzung garantiert hierbei die *Identität* des Subjekts mit sich selbst, während die Subjekt-Objekt-Struktur dagegen die *Epistemizität* (den Wissenscharakter) des Selbstbewußtseins sicherstellen soll. Dieter Henrich hat den Zirkel – den die Fichtesche Theorie des Ich umgehen will und in den sie doch zurückfällt – analysiert und in der Folge versucht, den Gedanken einer wissenden Selbstbeziehung des Subjekts weiterzudenken.[143] Ernst Tugendhat dagegen entwickelte in seiner dritten und vierten Heidelberger Vorlesung von 1974/75 eine sehr luzide sprachanalytische Kritik sowohl der Probleme klassischer Subjektivitätsstheorien als auch der Versuche der Heidelberger Schule, diese traditionellen Ansätze fortzuführen. Über die zentrale, auch von Henrich untersuchte, Schwierigkeit der „Reflexionstheorie" des Ich sagt Tugendhat:

> Das Selbstbewußtsein soll doch ein Ichbewußtsein sein. Ein Ich aber, so hörten wir, soll etwas nur dann sein, wenn es die Struktur der Identität von Wissendem und Gewußtem hat. Wenn nun aber das Selbstbewußtsein gemäß der Reflexionstheorie sich in einem Sichzurückwenden auf sich selbst vollziehen soll, dann wird erst im Akt dieser Rückwendung jene Identität von Wissendem und Gewußtem hergestellt. Andererseits soll das Subjekt, auf das sich der Akt zurückwendet, bereits ein Ich sein. Der Akt soll also einerseits, indem er sich zurückwendet, das Ich vorstellen, und andererseits konstituiert sich das Ich gemäß dem Begriff vom Ich erst in diesem Akt. Daraus ergibt sich, wie Henrich zeigt, ein Zirkel. Indem die Reflexionstheorie ein bereits vorhandenes Subjekt

---

einen unwillkürlichen, automatischen Mechanismus (die *réflexion virtuelle*) annimmt, der jeden mentalen Akt begleitet und ihn zu einem *bewußten* macht. Malebranche ging davon aus, daß Ideen extramentale Entitäten (nämlich solche im Geiste Gottes) sind, die dem Geist somit sehr nahe sind, aber dennoch außerhalb von ihm liegen. Sie sind Objekte des Geistes, *durch* die er die Welt erblickt (Eine ganz ähnliche Metapher finden wir heute bei van Gulick. Vgl. Van Gulick 1988: 178; siehe auch Fußnote 37). Als nach dem Modell einer „inneren Sinneswahrnehmung" gedachte Objekte sind sie das eigentliche *être représentatif*. Die echten Sinneswahrnehmungen dagegen besitzen keine Repräsentationsfunktion, sie stellen Modifikationen des Geistes dar, die den Ideen eine zusätzliche sinnliche Qualität verleihen. Für Malebranche muß der Geist also immer in direktem Kontakt mit den mediatisierenden mentalen Objekten sein, auch das Wissen um unseren eigenen Körper ist durch sie vermittelt. Arnauld – der den *anderen* Aspekt des cartesianischen Idee-Begriffs verteidigte – warf Malebranche vor, den Geist in einem Palast aus Ideen einzusperren und von der Welt auszuschliessen. Er selbst dagegen spricht von formalen Repräsentationen und bezieht sich damit auf mentale Akte. Für ihn sind alle Modifikationen der Seele Träger von repräsentationalem Gehalt, und das Zustandekommen dieser Modifikationen wird (wie auch die sinnliche Wahrnehmung) als Aktivität der Seele und als unvermitteltes, *direktes Bewußtsein* interpretiert. Arnauld interessiert sich nicht für Sinnestäuschungen und Fehlrepräsentationen, weil er repräsentationale Entitäten (wie vor ihm Ockham) für überflüssiges, die Homunculi des „inneren Sehens" erzeugendes Beiwerk hält: Er ist *direkter Realist* (vgl. Scheerer 1990b: 29).

[143] Vgl. Henrich 1967, 1970.

> voraussetzt, setzt sie schon das voraus, was sich in Wirklichkeit erst in der Beziehung auf sich konstituieren soll.[144]

Könnte eine naturalistische Theorie des Bewußtseins sich in einen ähnlichen Zirkel verstricken wie die idealistischen Theorien des *Selbstbewußtseins*? Man könnte dies meinen, wenn man „Meta-Repräsentation" als einen „reinen" – also im Grunde: nicht-natürlichen – Vorgang der Wissensgewinnung interpretiert, der sich nun selbst zum Objekt wird. Diese Gefahr besteht im begrifflichen Rahmen einer naturalistischen Theorie des Geistes jedoch deshalb nicht, weil **erstens** „Repräsentation" und „Meta-Repräsentation" als distinkte Prozesse gedacht werden, die räumlich und zeitlich getrennt im Gehirn eines biologischen Organismus ablaufen. Das Material der empirischen Forschung legt eine solche Sichtweise nahe. Die empirische Forschung zeigt uns nämlich, daß es ein grober phänomenologischer Fehlschluß wäre, von der *erlebten* „Instantanität" oder „Spontaneität" phänomenalen Bewußtseins auf eine Unmittelbarkeit der ihm zugrundeliegenden höherstufigen Repräsentationsvorgänge im Gehirn zu schliessen. Repräsentat und Repräsentandum sind dem naturalistischen Modell zufolge nicht identisch.

Phänomenales Bewußtsein ist darüber hinaus jedoch **zweitens** ein epistemisches Phänomen: In ihm wird etwas gewußt. Das „Gewußte" ist jedoch keine propositionale Wahrheit und es liegt – ganz im Gegensatz zum klassischen „Reflexionsmodell" – auch keine Identität von Subjekt und Objekt vor. Denn der epistemische Agent (das „Subjekt metarepräsentationaler Erkenntnis") ist das jeweilige System als Ganzes. Das, was erkannt wird (das „Objekt metarepräsentationaler Erkenntnis"), sind gewisse innere Abbildungsvorgänge, die zum Beispiel im Gehirn eines Menschen ablaufen. Dadurch, daß sie von höherstufigen Abbildungsvorgängen noch einmal erfaßt werden, entsteht eine neue psychologische Eigenschaft: Bewußtheit. Diese neue psychologische Eigenschaft können wir besser verstehen, indem wir uns um überzeugende Instantiierungserklärungen[145] bemühen. Solche Erklärungen werden eine gründliche Systemanalyse mit einer Untersuchung der funktionalen Zustände, die zu der speziellen repräsentationalen Struktur des Systems führen, verbinden. Wir finden bei **MRT** also nicht das von den Kritikern der traditionellen Denkfigur des „Reflexionsmodells" diagnostizierte „Ineinanderschieben der Wissens- und der Identitätsrelation".[146] Im folgenden Kapitel wird deutlich werden, daß dies auch für das naturalistische Modell des Selbstbewußtseins gilt, weil die zugrundeliegende Denkfigur nicht die am optischen Modell des geistigen Auges orien-

---

[144] Vgl. Tugendhat 1989 (1979): 62. – Vgl. ebenfalls Frank 1991, Kapitel I.1, II.2 und II.3. Frank weist darauf hin, daß im Gedanken des Subjekts als eines Sich-selbst-Vorstellens „die frühabendländische Metapher vom Bewußtsein als geistigem Schauen ihren Einstand feiert" (1991: 24) und daß das widersprüchliche Reflexionsmodell des Selbstbewußtseins immer noch der Figur des Gegenstandsbewußtseins verhaftet bleibt.

[145] Vgl. etwa Cummins 1983, Eimer 1990.

[146] Vgl. etwa Frank 1991 und Tugendhat 1979.

tierte „Reflexion“ ist, sondern die unter dem Informationsverarbeitungsansatz entwickelte Idee der „internen Selbstmodellierung“.

Kehren wir noch einmal kurz zurück zur logischen Struktur der Metarepräsentationsbeziehung, so wie ich sie oben dargestellt habe. Die philosophisch vielleicht interessanteste Einsicht versteckt sich in der letzten Zeile: Mentale *Meta-Simulation* kann es im strengen Sinne nicht geben. Um kontrafaktische mentale Repräsentate, Simulate, oder sogar Präsentate in Form einer bewußten Simulation intern darzustellen, muß ein physikalisch realisiertes Repräsentationssystem diese automatisch in gewissem Umfang in sich aktivieren – und damit wird aus der Simulation eine Repräsentation, weil nun ihr Gegenstand intern gegeben ist. Da Meta-Repräsentation in der Einbettung eines mentalen Modells in ein anderes besteht[147], *enthält* das Bewußtsein erzeugende Meta-Repräsentat (das bewußte phänomenale Weltmodell) automatisch alle von ihm dargestellten mentalen Zustände – weil sie dank ihrer relationalen Struktur in *seine* relationale Struktur eingebunden werden konnten. Dies mag etwas abstrakt klingen, läßt sich aber durch einen Blick auf die Alltagspsychologie verdeutlichen. Wenn ich einem Tagtraum nachhänge (das Gehirn also eine mentale Simulation durchführt) und dabei auch über meine vergangenen, zukünftigen oder möglichen Gedanken und Gefühle nachdenke, dann werden diese Gedanken und Gefühle automatisch – wenn auch vielleicht nicht in ihrem vollen qualitativen Gehalt – mitaktiviert, sie werden zu Bewußtseinsinhalten und damit zu aktuellen mentalen Zuständen. Läßt man Magie und Psychokinese einmal beiseite, dann zeigt sich andererseits, daß systemexterne Zustände der Welt nicht „aktiviert“ werden können, bloß weil wir sie mental simulieren. Manche der internen Zustände eines Systems werden jedoch durch Meta-Repräsentation automatisch mitaktiviert, dadurch werden sie nicht nur zu Bewußtseinsinhalten, sondern erscheinen intuitiv als besonders verläßlich und gewiß.[148]

In Analogie zu den Überlegungen des vorangegangenen Abschnitts kann man nun sagen: Die Homogenität des phänomenalen Bewußtseins ist eine Illusion[149], die durch einen niedrigen Auflösungsgrad derjenigen Funktion bedingt ist, die mentale Repräsentate zu bewußten macht. Angenommen, es gibt eine metarepräsentierende Funktion, die durch „reflexive“ Operationen auf bestimmten mentalen Repräsentaten höherstufige Repräsentate erzeugt, die *als Repräsentate* selbst unbewußt bleiben, aber ihren Reprä-

[147] Vgl. dazu das **„Pflichtenheft für mentale Repräsentate“** und den folgenden Abschnitt.

[148] Auf die Frage nach der Selbstgewißheit des Subjekts komme ich in Kapitel 3 zurück.

[149] Interessant erscheint in diesem Zusammenhang vielleicht auch, daß es im außereuropäischen Raum seit langem philosophische Psychologien gibt, die die Auflösbarkeit der Homogenitäts-Illusion – sowohl was die phänomenale Welt als Ganze, als auch was die Einheit des Erlebnissubjekts angeht – durch verstärkte Introspektion lehren. Zum Beispiel zielt eine Reihe buddhistischer Meditationstechniken genau darauf ab, durch systematische Erhöhung innerer Aufmerksamkeit jene Funktionen, die die Illusion der Einheitlichkeit des Selbst erzeugen, zu Repräsentanda zu machen und so die Mechanizität und Fragmentiertheit der Grundlagen phänomenalen Bewußtseins *in dieses selbst* als Repräsentat einzuführen. Vgl. Nyanaponika 1975.

sentanda die fragliche Qualität verleihen. Dann wäre **erstens** Bewußtsein aus der Perspektive der Wissenschaft keine instantane Qualität mehr, weil der es erzeugende neurobiologische Prozeß Zeit benötigt. Dieser Umstand macht es für das Gehirn notwendig, seine Informationsverarbeitung zeitlich zu quanteln[150] und gewisse Reize etwa zu antedatieren[151], um zu einem homogenen, multimodalen Modell des fraglichen Repräsentandums zu gelangen. Das bedeutet, daß das subjektive Jetzt den vielen objektiven Ereignissen, durch die es konstituiert wird, hinterherhinkt: Bewußtes Erleben ist im Grunde eine Form von Erinnerung. **Zweitens** könnte diese Funktion Diskontinuitäten, zeitliche Brüche oder Ambiguitäten ausfiltern, indem sie mehrere Input-Repräsentanda der tieferliegenden Repräsentationsebenen zusammenfaßt[152] zum Output von wesentlich weniger Repräsentaten auf der Ebene phänomenalen Bewußtseins. Vollständige Homogenität wäre aber nur zu erreichen, wenn die gesamte Information in ein *einziges* Repräsentat zusammengeführt würde. Dieses Repräsentat, das aktuelle bewußte Realitätsmodell des Systems, wäre somit eine jeweils gerade aktive Datenstruktur mit sehr spezifischen abstrakten Eigenschaften.

Deshalb kann man diesen Gedanken möglicherweise auch vom Begriff des „Formats" her zu entwickeln versuchen. Wir haben bereits gesehen, daß mentale Repräsentate in verschiedenen Formaten auftreten. Eines der philosophisch interessantesten Charakteristika der Ebene subjektiven Erlebens ist dabei Multimodalität – das Vorhandensein unterschiedlicher Repräsentatformate – bei gleichzeitiger Homogenität. Wenn es prinzipiell möglich ist, die phänomenale Ebene als einen speziellen Sonderfall oder Teilbereich interner Informationsverarbeitung zu erklären, dann muß es eine neuroinformatische Antwort auf die Frage geben, wie die verschiedenen Formate bruchlos integriert werden können zu der phänomenalen Repräsentation *einer* Welt, in der dennoch verschiedene Wissens- und Wahrnehmungsquellen unterschieden werden können. Vielleicht stellt die bewußte Ebene ein eigenes Repräsentationsformat dar. Das würde bedeuten, daß die fragliche Homogenität eben genau aus der Tatsache resultiert, daß wir es – beim Gehalt von phänomenalem Bewußtseins *als Ganzem* – überhaupt nur mit *einem* Repräsentat zu tun haben. Dieses Repräsentat besitzt einen eigenen Set abstrakter Eigenschaften, eben sein Format. Das Format könnte letztlich das sein, was wir als die verbal so schwer zu beschreibende „Feldqualität" unseres phänomenalen Raums erleben. Man kann vermuten, daß dieser Set von abstrakten Eigenschaften innerhalb

[150] Vgl. etwa Libet 1979, Pöppel 1985, 1989.

[151] Vgl. Fußnote 71.

[152] Wenn diese empirische Spekulation in die richtige Richtung geht, dann bestimmt das Auflösungsvermögen der Metarepräsentationsfunktion (über die „Körnung" des von ihr erzeugten Repräsentats; vgl. Sellars 1963, Lycan 1987) die Größe der kleinsten phänomenalen Einheiten. Das durch die Kapazität der neurobiologischen „Wetware" begrenzte phänomenale Zeitbewußtsein besteht zum Beispiel aus subjektiven „Augenblicken" von maximal 3 Sekunden. Innerhalb der durch das „Gegenwartsfenster" der neuronalen Mechanismen determinierten inneren Jetzte herrscht bezüglich des Zeiterlebens dann genau die fragliche temporale Homogenität.

unserer Spezies relativ konstant ist, muß aber gleichzeitig erkennen, daß andere Metarepräsentationssysteme ganz andere Realitätsmodelle in anderen Formaten konstruieren könnten. Deshalb hätten sie auch einen anderen Typ von „Bewußtheit", sie würden eine andere Position im Möglichkeitsraum phänomenaler Systeme einnehmen. „Bewußtheit" wäre somit letztlich eine Menge von abstrakten Eigenschaften einer neuronalen Datenstruktur, die wiederum supervenient gegenüber den repräsentationalen Eigenschaften derselben ist. Das heißt: Wenn alle repräsentationalen Eigenschaften des betreffenden Makro-Repräsentats feststehen, dann stehen auch alle seine abstrakten Eigenschaften fest und damit auch, welchen Typ von Bewußtheit (welche *psychologischen* Eigenschaften) das betreffende System als Ganzes instantiiert.

Der dieses Makro-Repräsentat erzeugende Vorgang selbst müßte *ex hypothesi* bewußtseinsextern bleiben. Wenn es die „höchsten" kognitiven Vorgänge sind, welche via Meta-Repräsentation die psychologische Eigenschaft der Bewußtheit generieren, dann werden genau diese Vorgänge immer unbewußt bleiben müssen. Denkt man also noch in einem hierarchischen Modell der Informationsverarbeitung und erklärt die phänomenale Ebene des Systems durch Meta-Repräsentation, dann wird diese Ebene immer eine „mittlere" Ebene sein – und nie die höchste: Bewußtsein ist eben gerade nicht die „Spitze des Eisbergs" oder das „Cartesianische Theater".[153] Trotzdem gäbe es eine gemeinsame Eigenschaft, die alle von der fraglichen Funktion erfaßten internen Repräsentanda verbände: Gehalt *eines* einheitlichen Repräsentats, eines einzigen umfassenden Subsymbols zu sein. Dieses Repräsentat wäre *das bewußte phänomenale Modell der Welt.* Ein solches umfassendstes Makro-Repräsentat könnte dynamisch sein und durch ständige Korrekturen in Echtzeit eine große inhaltliche Variabilität aufweisen. Trotzdem wäre es für uns *qua Erlebnissubjekte* unhintergehbar, da es auf seiner Komplexitätsstufe das einzige wäre und introspektiv nicht mehr mit anderen Repräsentaten desselben informationellen Reichtums verglichen werden könnte. Wenn diese Spekulationen in die richtige Richtung gehen, dann muß es ein funktionales und/oder strukturelles Äquivalent dieses umfassendsten Makro-Repräsentats im Gehirn geben. Es müßte prinzipiell auch möglich sein, die es generierende neuronale Funktion zu beschreiben – diejenige Funktion also, die aus der Gesamtmenge der zu einem gegebenen Zeitpunkt im System aktivierten Simulate, Repräsentate und Präsentate das phänomenale Weltmodell herausgreift und dadurch erzeugt.

Ich hoffe mit diesen knappen Bemerkungen angedeutet zu haben, wie auch der Bewußtheits-Aspekt von Subjektivität als Eigenschaft natürlich entstandener Repräsentationssysteme verstanden werden könnte. Mehr läßt sich seriöserweise beim derzeitigen Wissensstand nicht sagen. Empirisch könnte man die oben angebotene Proto-Hypothese anreichern und präzisieren, indem man solche Fälle untersucht, in denen Bewußtsein sich

[153] Vgl. Dennett 1991: Kapitel 5.

spaltet oder heterogen wird.[154] Von theoretischem Interesse scheint außerdem die Frage zu sein, ob (wie etwa Flohr denkt) ein kritisches Maß an funktionaler Komplexität *automatisch* zu Metarepräsentation und interner Selbstdarstellung führt[155], oder ob es noch weitere spezifische Hintergrundbedingungen und Mechanismen der Eigenschaftsbindung für Bewußtsein beim Menschen (und physikalischen Repräsentationssystemen im allgemeinen) gibt. Lassen Sie uns jedoch vorerst untersuchen, welches Bild der für subjektives, phänomenales Bewußtsein verantwortlichen mentalen Repräsentate sich aus den bisher angestellten Überlegungen ergibt.

2.2 Mentale Modelle: Die abstrakten Werkzeuge biologischer Systeme

Wir haben in den vorangegangenen Abschnitten verschiedene Aspekte von mentalem Gehalt betrachtet, die für eine Theorie der Subjektivität von Interesse sind. Ich habe diese Aspekte als Eigenschaften analysiert, die

---

[154] Edoardo Bisiach unterscheidet drei fundamentale Bewußtseinsbegriffe: C1 (das strikt private, erlebte Bewußtsein aus der Innenperspektive des psychologischen Subjekts), C2 (der Zugang von Teilen oder Prozessen eines Systems zu anderen seiner Teile oder Prozesse; also der klassische funktionalistische Ansatz von Bewußtsein als dem *monitoring of internal representations* [1988: 104]) und C3 (die Hypostasierung von C1 zu nicht-physikalischen Entitäten). Das Problem besteht in einer Abbildung der Zustände von C2 (dem legitimen Objekt wissenschaftlicher Untersuchungen) auf C1. Bisiach hat an Beispielen gezeigt, welche theoretischen Konfusionen allein durch den äquivoken Gebrauch von C1, C2 und C3 entstehen und anhand von Beispielen die Fraktionierung von C2 in pathologischen Situationen (aber auch bezüglich der zeitlichen Struktur von Bewußtsein) demonstriert, also die Möglichkeit des Zerfalls und der funktionalen Dissoziierbarkeit von Metarepräsentationsfunktionen im menschlichen Gehirn. Solche empirischen Befunde (wie zum Beispiel Anosognosien, also die Unbewußtheit von Bewußtheitsdefiziten) sind für die uns hier beschäftigenden theoretischen Fragen deswegen von Interesse, weil sie einerseits die Modularität des computationalen Bewußtseins in Kontrast zu der subjektiv erlebten Einheitlichkeit auch pathologischer Zustände setzen. Andererseits aber zeigen sie, wie stark phänomenales C1-Bewußtsein „von unten" determiniert wird. *„Thus a patient may be selectively unaware of his blindness, or deafness, or hemiplegia if the different brain lesions which are responsible for one or the other of these disorders impair that level. At the same time, the patient may be fully aware of a concurrent disorder (e.g. of dysphasia) if this is due to the description of more peripheral processing stages. This fractionation of anosognosia into function-specific forms shows that monitoring of inner activity is not accomplished in the nervous system by a unitary, superordinate entity watching the workings of its slave mechanisms and able to detect faults in their operations as soon as they occur. Monitoring of inner activity is rather to be viewed as a function distributed across the different analysers to which it refers (. . .). Indeed, from a neuropsychological perspective, C2 is more properly viewed as a collective name for a bundle of dissociable processes, conforming to Gazzaniga's ‚sociological' concept of consciousness (. . .)."* Vgl. Bisiach 1988: 108f.

[155] Das ist eine wichtige Frage in Zusammenhang mit der Flohrschen Hypothese: Organisieren sich metarepräsentationale Strukturen automatisch selbst, wenn das System ein gewisses Komplexitätsniveau erreicht hat, oder müssen weitere repräsentationale Bedingungen erfüllt sein, bevor phänomenales Bewußtsein entstehen kann? Vgl. Flohr 1991, 1992a,b; zur Selbstorganisation kognitiver Strukturen vgl. Singer 1989, 1990.

mentale Repräsentate mindestens besitzen müssen, um den philosophisch interessanten psychischen Phänomenen zugrunde liegen zu können. Die Eigenschaften dieses theoretischen Prototyps eines phänomenalen Repräsentats waren:

- *wechselseitige Einbettbarkeit* (Formatintegration/Multimodalität)
- *Analogizität* (Darstellung von Intensitäten und Kontinua)
- *Dynamizität* und *Simulationsfähigkeit* (Inputunabhängigkeit)
- *Präsentation* (Signalfunktion ohne automatische Output-Erzeugung)
- potentielle *Meta-Repräsentierbarkeit* („Bewußtheit", „semantische Transparenz")

Dazu habe ich begrifflich zwischen Präsentaten, Repräsentaten, Simulaten und Meta-Repräsentaten unterschieden. Aber natürlich ist unser phänomenales Bewußtsein wesentlich homogener, als diese theoretischen Unterscheidungen vielleicht suggerieren: Das Buch, das Sie momentan in Händen halten, ist Ihnen mit dem qualitativen Gehalt seiner Farbigkeit, seines Gewichts, seiner taktil erlebbaren Oberflächentextur, mit seinem wirklichen und möglichen Inhalt gegeben – und zwar *bewußt*. Der Präsentationsaspekt und die phänomenale Tatsache, daß Sie sich des Buches bewußt sind, stellen aber keine deutlich unterscheidbaren Elemente ihres Bucherlebnisses dar. Wenn wir die bruchlose Natur der phänomenalen Ebene verstehen wollen, müssen wir also nach einer Entität suchen, die alle in den vorangegangenen Abschnitten besprochenen Aspekte mentaler Repräsentate potentiell in sich vereint.

In Anlehnung an Arbeiten von Colin McGinn[156] und Philip Johnson-Laird[157] werde ich jetzt den bereits in groben Umrissen entwickelten Begriff des „mentalen Modells" einsetzen, um diejenigen Eigenschaften näher zu beleuchten, die mentale Repräsentate besitzen müssen, wenn sie die Instantiierung subjektiven, phänomenalen Bewußtseins unterstützen sollen. Das Gehirn ist ein hochkomplexes Organ und man kann davon ausgehen, daß es eine Vielzahl von Repräsentattypen intern einsetzt. Unterschiedliche theoretische Fragestellungen werden dabei die Aufmerksamkeit auf verschiedene mögliche „repräsentationale Kandidaten" und innere Strukturen lenken. Wenn man an der Frage interessiert ist, wie *subjektives* Bewußtsein aus mentaler Repräsentation resultieren kann, muß man Antworten auf Fragen der folgenden Art zu geben versuchen: *Wie könnten die internen Strukturen beschaffen sein, die den fraglichen, die Subjektivität phänomenalen Bewußtseins konstituierenden psychischen Phänomenen zugrundeliegen? Welche Eigenschaften müssen mentale Repräsentate besitzen, um die*

[156] Vgl. McGinn 1989: 172ff.
[157] Vgl. Johnson-Laird 1983, 1989.

zu den Phänomenen führenden Operationen zu unterstützen? Aus diesem Grund werde ich jetzt versuchen, den Begriff des mentalen Modells anzureichern, indem ich einige der oben bereits angegebenen Kriterien in Form von Thesen vorstelle und näher erläutere.

(1) Mentale Modelle sind multimodale Datenstrukturen.

Die Inhalte unseres phänomenalen Bewußtseins zeichnen sich dadurch aus, daß Information aus verschiedenen Quellen bruchlos zu einheitlichen mentalen Modellen verschmolzen wird. Wenn Sie dieses Buch simultan fühlen *und* sehen, dann erleben Sie es als *eines* – als singuläres Objekt in der Außenwelt, das Ihnen durch zwei verschiedene Sinne gegeben ist. Das mentale Modell des Buches enthält nämlich noch die Information[158], daß es auf mindestens zwei unterschiedliche Weisen gegeben ist: Visuelle und taktile Eigenschaften werden zu einer erlebnismäßig einheitlichen Struktur verbunden.[159] Es ist dies der Prozeß, den ich weiter oben als Formatintegration bezeichnet habe. Durch die Verschmelzung zweier interner Repräsentate unterschiedlicher Genese und unterschiedlichen Formats entsteht eine höherstufige mentale Struktur. Diese einheitliche Struktur ist das mentale Modell des Buches.

---

[158] Diese Information wird dadurch dargestellt, daß ein mentales Modell mit zwei Präsentatformaten unterlegt wird. Vgl. Abschnitt 2.1.3.

[159] Die subjektive Einheitlichkeit muß dann nicht mehr *von oben* erklärt werden, wenn es eine gute neurowissenschaftliche Theorie derjenigen Leistung gibt, die in der Fachterminologie der Hirnforschung als *feature binding* bezeichnet wird: Die Fusionierung verschiedener Objekteigenschaften – wie Kanten, Bewegungen oder Farben (die aber nachweislich durch weit auseinanderliegende Neuronenverbände im Gehirn repräsentiert werden) – zu einer einheitlichen internen Struktur. Wolf Singer und seine Mitarbeiter haben entdeckt, daß weit verteilte Neuronen im Gehirn von Katzen, die auf von ein und demselben visuell präsentierten Objekt ausgehende Stimuli reagieren, synchron mit einer Frequenz von 30 bis 80 Hertz zu oszillieren beginnen. Daß die Bindung visueller Eigenschaften durch sehr kurze Synchronisierungen verteilter Aktivitätsmuster geleistet werden könnte, hatte Christoph von der Malsburg schon 1981 vorgeschlagen. Diese neuen Entdeckungen zeigen nun, daß sich tatsächlich für sehr kurze Perioden von weniger als einer halben Sekunde solche Oszillationen des Feldpotentials mit etwa 40 Hz etablieren. Es ist noch zu früh, um zu folgern, daß dies die neurobiologischen Korrelate von phänomenalen Eigenschaften wie „Aufmerksamkeit" oder „Bewußtheit" sind – schließlich könnten wir hier auch einfach mit afunktionalen Epiphänomenen konfrontiert sein, wie sie in komplexen Systemen immer zu finden sind. Interessant ist an diesen Prozessen allerdings, daß sie sich bei der Selbstorganisation an den klassischen Gestaltkriterien wie etwa Nachbarschaft, Ähnlichkeit, Bewegungskontinuität usw. orientieren und auch, daß sie in etwa die selbe Zeit beanspruchen, in der die Aufmerksamkeit einer Person von einem Objekt zum anderen springt. *„One of the features that makes the 40-hertz oscillations attractive as a mediator of visual awareness, . . ., is that their time scale corresponds with that of attention flitting from one object to another. The neurons typically stay phase-locked for several hundred milliseconds, which would allow them to make and break their liaisons in roughly the same period that a person's attention moves from one subject to the next. As different subjects compete for attention, different sets of neurons may set up oscillations, . . . One wins momentarily, and attention is briefly focussed. Then that oscillation fatigues and attention is directed elsewhere."* (Barinaga 1990: 857). Ein hochauflösender zeitlicher Code um Relationen auszudrücken ist deshalb auch philosophisch interessant, weil er auf theoretischer Ebene verständlich machen könnte, wie das Gehirn aus Eigenschaftsmengen *Objekte* bilden und diese durch

Der theoretisch entscheidende Punkt liegt in der Tatsache, daß wir bald über befriedigende naturalistische Erklärungen darüber verfügen werden, wie die Multimodalität gewisser komplexer, von biologischen Gehirnen erzeugter Datenstrukturen das Resultat von neuronaler Informationsverarbeitung sein kann. In Abwesenheit solcher Erklärungen war es nur zu verständlich, daß Philosophen des Geistes phänomenale Individuen[160] und intentionale Objekte postuliert haben: Wie sonst sollte die bruchlose Integration von unterschiedlichen Sinneseindrücken, Erinnerungen usw. zu einheitlichen Erlebnissen zu erklären sein? Wir sind aber heute in der glücklichen Lage, an dieser Stelle keine nicht-physischen Entitäten mehr postulieren zu müssen, die uns die multimodale Bindung von Eigenschaften zu verstehen helfen. Das konnektionistische Erklärungsmodell bietet Hoffnung auf eine zukünftige neuroinformatische Erklärung dafür, wie in verschiedenen Formaten vorliegende, durch Aktivierungsvektoren beschreibbare interne Repräsentate miteinander verknüpft werden können, obwohl sie in unterschiedlichen Subsystemen des Gehirns erzeugt wurden. Deswegen sind wir nicht mehr gezwungen, von einer Integration oder Eigenschaftsbindung „von oben" auszugehen. Wenn man so will, sind mentale Modelle *naturalisierte intentionale Objekte* – wobei der sie erzeugende Prozess des Informationsflusses in die entgegengesetzte Richtung zeigt wie der Pfeil der Intentionalität.

In pathologischen Fällen ist es möglich, daß diese Verknüpfungsleistungen nicht mehr erbracht werden können. Bei manchen Arten von Hirnverletzungen kann es zu funktionalen Dissoziationen[161] kommen, welche dann *phänomenale* Entkoppelungen zu Folge haben. Ray Jackendoff zitiert einen Bericht von Anthony Marcel[162] über einen Patienten, der, wenn angesprochen, die gesehenen Lippenbewegungen seines Gegenübers und sein auditorisches Hörerlebnis nicht mehr integrieren konnte. Dieser Patient beschrieb seinen phänomenalen Zustand als von zwei vollkommen unterschiedlichen Ereignisketten geprägt und verglich ihn mit einem alten Ton-

---

immer grösser werdende „Zeitfenster" ineinander einbetten kann. Da sich bei verändertem Input dieselben Zellen zu immer neuen Verbänden zusammenschliessen können, ist diese Hypothese nicht nur sparsam in strukturellen Annahmen, sondern erklärt auch die Flexibilität und die Plastizität der so entstehenden repräsentationalen Zustände. Vgl. Barinaga 1990, Crick/Koch 1990a, b, Engel *et al.* 1991a, b, c, 1992a, b, c, Gray *et al.* 1989, Kreiter/Singer 1992, Singer 1989, 1990, 1993, 1995, von der Malsburg 1981, 1986.

[160] Vgl. Lycan 1987, Kapitel 8.

[161] Das Studium solcher Dissoziationen ist eine der wichtigsten Quellen von Einsichten bezüglich der funktionalen Architektur, die phänomenalen Zuständen zugrundeliegt, und hat bereits viele der klassischen *a-priori*-Annahmen über das Wesen des Bewußtseins falsifiziert. *„. . ., the most important findings in neuropsychology concern what dissociations are possible, regardless of how often they occur. Even a single demonstration of a highly specific loss, in the absence of other deficits, provides a rich source of inferences about the organization of processing in the nervous system; and instances of ‚double dissociation' of two different kinds of capacities allow relatively direct inferences to be drawn about potentially independent processes (Weiskrantz 1968).*" Vgl. Weiskrantz 1988: 188.

[162] Vgl. Marcel 1983: 292f.

film, bei dem eine Tonspur mitläuft, die in keinerlei Beziehung zu der gerade visuell erlebten Konversation steht. Jackendoff schreibt über die Rolle unbewußter kognitiver Operationen für die Vereinheitlichung separater Bewußtseinsmodalitäten:

> . . . : How is it that entities detected in multiple modalities can be experienced as unified? For instance, when I look at something and handle it at the same time, how can I experience it as the *same* object, if my awareness is disunified into visual and haptic modalities?
> The answer comes from the character of processing. The visual and haptic representations that support awareness of the object are each in registration with 3D model and conceptual structures that encode the shape, identity, and category of the object. If it so happens that they are in registrations with the *same* 3D model and conceptual structure, then the two modalities will be understood and experienced as simultaneous manifestations of the same object. In other words, understanding again parses out experience, this time across modalities.[163]

Aus Untersuchungen neuronaler Informationsverarbeitung und dem Studium von Diskonnektions-Syndromen ergibt sich allmählich ein überzeugendes wissenschaftliches Bild der Integration multimodaler Information im phänomenalen Bewußtsein. Aber kann dieses Bild auch *intuitiv* befriedigend sein? Schließlich erleben wir weder Objekte in unserer Umgebung noch subjektive, innere Zustände jemals *als* mentale Modelle. Dies gelingt uns auch dann nicht, wenn wir die hier skizzierte Theorie des Geistes für richtig halten: Die betreffenden Operationen sind kognitiv nicht penetrabel.[164] Das bedeutet, daß die Konstruktion mentaler Modelle weitgehend auf unbewußten Prozessen beruht, die absichtlich eingeleiteten mentalen Operationen nicht zugänglich sind. Es ist aber gerade jene homogene Verknüpfung der durch unterschiedliche Transduktoren und Module abgebildeten Eigenschaften von Repräsentanda, die uns schon vor aller theoretischen Reflexion zu naiven Realisten werden läßt. Der naive Realismus ist nämlich genau die Theorie, die den phänomenalen Gehalt unseres mentalen Modells der Welt – unter den Standardbedingungen nicht-pathologischer Wachzustände – erläutert. Das soll heissen: Gerade *weil* das mentale Modell dieses Buches die Trennung (etwa der haptischen und der visuellen) Modalitäten bei gleichzeitiger Integration der gelieferten Information aufrechterhält, entsteht die Illusion, daß wir es *direkt* mit einem externen Objekt zu tun haben und nicht mit einem komplexen mentalen Repräsentat. Dadurch, daß *eine* homogene mentale Struktur simultan durch *verschiedene* Repräsentationssubsysteme – zum Beispiel durch mit taktiler und mit visueller Information arbeitenden – gegeben ist, entsteht auf der phänomenalen Ebene der Eindruck, es mit einem auf verschiedene Weisen gegebenen Individuum zu tun zu haben.

---

[163] Vgl. Jackendoff 1987: 300f.

[164] Der Begriff der *cognitive impenetrability* stammt von Zenon Pylyshyn. Vgl. Pylyshyn 1980, 1984.

Dieser Eindruck einer Konfrontation mit phänomenalen Individuen liegt vielen unserer cartesianischen Intuitionen zugrunde, behaupte ich. Wir können die Eigenschaft der Multimodalität unserer mentalen Repräsentate nicht als konstruiert erleben, weil die zugrundeliegenden Prozesse der Formatintegration auf unbewußten (= nicht meta-repräsentationsfähigen) neuronalen Mechanismen zu beruhen scheinen. Als psychologische Subjekte sind wir mit komplexen mentalen Repräsentaten konfrontiert, die sowohl erlebnismäßig unfragmentiert, als auch ganz offensichtlich durch verschiedene Informationsquellen gegeben sind. Sind diese komplexen Repräsentate Elemente unseres Modells der Welt, dann gewinnen sie durch eben diese charakteristische Kombination die Qualität der *Objekthaftigkeit*. Sind sie Elemente unseres mentalen Modells unserer *selbst*, dann erhalten sie dadurch den ebenso unhintergehbaren Charakter der phänomenalen *Meinigkeit*.

Die Multimodalität komplexer mentaler Repräsentate ist eine wichtige Ursache für das philosophisch sehr interessante Phänomen, daß wir sie nicht als solche erkennen können. Auf diese Weise ist nämlich ein sehr einfacher und direkter Realismus in unser vorphilosophisches Welterleben eingebaut. Dieser unserem im Wachzustand von unseren Gehirnen konstruierten mentalen Modell der Wirklichkeit inhärierende intuitive Realismus ist eines der wesentlichen Hindernisse bei theoretischen Bemühungen bezüglich des Wesens geistiger Prozesse: Für uns alle ist die Versuchung sehr groß, einfach immer kurz hinzuschauen.[165]

(2) Mentale Modelle sind analoge Datenstrukturen.

Die Typen von internen Repräsentaten, die ein informationsverarbeitendes System zur Wissensgewinnung einsetzt, bestimmen wesentlich seinen Typ von Intentionalität, die Art des „Auf-die-Welt-Gerichtetseins" des Systems. Sie bestimmen *wie* Information verarbeitet und intern dargestellt werden kann. Außerdem haben sie einen direkten Einfluß auf die psychologischen Eigenschaften, die das System mit ihrer Hilfe instantiiert. Ich werde in Abschnitt 2.2.1 einige Bemerkungen über den Unterschied zwischen „digitalen" und „analogen" Codes als epistemischen Instrumenten machen. An dieser Stelle ist nur der Hinweis wichtig, daß analoge mentale Repräsentate empirisch besser belegte und explanatorisch wesentlich potentere Entitäten sind als linguistische interne Symbolketten à la Fodor – zumindest, wenn wir phänomenales Bewußtsein erklären wollen. Analoge Strukturen weisen eine Isomorphie zu ihren Repräsentanda auf – sie *ähneln* ihnen in ganz bestimmten Hinsichten. Das bedeutet: Von den Eigenschaften der Repräsentate sind Rückschlüsse auf die Eigenschaften der Repräsentanda möglich. Für digitale Repräsentate gilt dies nicht.[166]

[165] Wir laufen deshalb immer Gefahr, zu Arnauldianern zu werden, weil wir beim introspektiven Hinschauen doch immer nur durch die Repräsentate hindurchschauen: Die Schnelligkeit der zugrundeliegenden neuronalen Mechanismen macht uns zu direkten Realisten.

[166] Digitale mentale Repräsentate hatten ihren ersten Auftritt in der abendländischen

In unserem Zusammenhang besteht das wichtigste Charakteristikum analoger Repräsentate in ihrer partiellen relationalen *Homomorphie* bezüglich ihrer Repräsentanda. Mentale Modelle bilden relationale Eigenschaften ab: Sie geben die relationale Struktur der von ihnen modellierten Gegenstände in der Welt auch durch ihre eigene relationale Struktur wieder. Eine eigene relationale Struktur besitzen mentale Modelle wiederum durch ihre Vernetzung mit anderen inneren Zuständen des Systems.[167] Der paradigmatische Fall relationaler Simulation dürfte die Konstruktion des durch visuellen Input dominierten[168] räumlichen Modells unserer Umgebung sein. In ihm werden räumliche Relationen wie etwa die „Nachbarschaft" einzelner Repräsentanda durch interne räumliche Relationen wiedergegeben.[169] Um ein mentales Modell eines räumlichen Prozesses oder Objekts zu generieren, muß also vom Gehirn eine Funktion realisiert werden, welche die Relationen zwischen Elementen oder Teilen des Prozesses bzw. Objekts systematisch auf Relationen zwischen Teilen der erzeugten inneren Struktur abbildet.

Wenn man ein mit Repräsentaten operierendes System als „analog" bezeichnet (Etwa, indem man sagt: *„Das menschliche Gehirn ist in Wirklichkeit ein System aus mehreren miteinander in Wechselwirkung stehenden Analogrechnern!"),* bleibt eine Ambiguität bestehen. Denn Analogizität kann also sowohl als Eigenschaft mentaler Modelle als auch als Eigenschaft der Abbildungsrelation zwischen Repräsentandum und Repräsentat analysieren. Ein analoge Repräsentate erzeugendes System realisiert also eine ganz bestimmte Abbildungsfunktion, die die inneren Gesetzmäßigkeiten und die relationale Struktur des Originals im Modell bewahrt. Von den Eigenschaften des Modells können Rückschlüsse auf Eigenschaften des Originals gezogen werden, weil die Operationen des Systems gewisse Isomorphismen erzeugen. In der Literatur werden unterschiedliche Isomorphismen diskutiert, zum Beispiel konkrete, physikalische Ismorphismen

---

Geistesgeschichte wahrscheinlich in der psychologistischen Logik der Spätscholastik und ihrer *oratio mentalis.* Hier fand wohl erstmals die Anwendung *konventioneller* Repräsentationssysteme als theoretisches Modell für natürliche Repräsentationssysteme ihren Eingang in die Philosophie des Geistes. Vgl. Scheerer 1990a: 16f.

167 Man kann annehmen, daß die Relationen zwischen mentalen Modellen realisiert sind als die Bedingungen ihrer Aktivierung durch das jeweilige physische System.

168 Es gibt auch ein akustisches und ein kinästhetisches Raumbild. Eine gute Analyse der an der Erzeugung des *visuellen* Realitätsmodells beteiligten mentalen Repräsentationsleistungen in Form von Instantiierungserklärungen gibt Martin Eimer. Vgl. Eimer 1990; Marr 1982.

169 Das bedeutet nicht, daß es auch eine Größenkovarianz zwischen Objekt und Repräsentat gibt. In der Tat scheint dies bei „somtatotopischen Karten" in biologischen Gehirnen nicht der Fall zu sein. Helge Ritter demonstriert die Entstehung einer solchen nachbarschaftserhaltenden Projektion durch die Computersimulation eines Selbstorganisationsprozesses auf neuronaler Ebene in Ritter 1988: 59ff. Vgl. dazu auch Ritter/Martinetz/Schulten 1990: Kapitel 7.

170 Über den Unterschied zwischen Isomorphismen erster und zweiter Ordnung (in der Konzeption von Shephard/Chipman 1970) schreibt Rehkämper: *„Bei einem Isomorphismus erster Ordnung werden sowohl einstellige als auch mehrstellige Relationen abgebildet,*

und funktionale Isomorphismen zweiter Ordnung.[170] Letztere sind für das Phänomen der mentalen Repräsentation am bedeutendsten. Auf empirischer Ebene gibt es gute Belege dafür, daß das menschliche Gehirn tatsächlich mit analogen Modellen operiert. Ein Beispiel sind die klassischen und auch von Philosophen vielzitierten Untersuchungen von Shephard und verschiedenen seiner Mitarbeiter[171] zur mentalen Rotation dreidimensionaler Objekte.[172]

Eine Frage, die sich in Anschluß an die Einsicht erhebt, daß es analoge Operationen sein müssen, die unserem phänomenalen Bewußtsein zugrundeliegen, lautet: Sind Menschen – als natürliche Repräsentationssysteme betrachtet – *Hybridsysteme*[173], die intern sowohl mit analogen als auch mit propositionalen Formaten arbeiten (d. h. verschiedene Algorithmenklassen realisieren), oder können auch die klassischen Beispiele der mentalen Operation mit propositionalem Gehalt unter mentale Modellierung subsumiert werden? Sind wir ausschließlich Modellgeneratoren oder – wenigstens partiell – auch *syntactic engines*? Man kann hier zwei mögliche Thesen unterschiedlicher Stärke unterscheiden:

Die Starke Analogizitäts-These (SAT):

Beim Menschen werden alle mentalen Phänomene durch die interne Manipulation analoger Strukturen erzeugt. Analoge Operationen sind alles, was wir brauchen, um die Genese von mentalem Gehalt zu verstehen.

Die Partielle Analogizitäts-These (PAT):

Ein Teil der mentalen Phänomene beim Menschen wird durch die interne Manipulation analoger Strukturen erzeugt. Manche psychologischen Eigenschaften und manche Formen mentalen Gehalts lassen sich nur über analoge

---

bei einem Isomorphismus zweiter Ordnung nur Relationen mit einer Stelligkeit > 2; die abbildenden Relationen haben dabei natürlich jeweils die Stelligkeit des Originals. Bei einem Isomorphismus erster Ordnung werden die Eigenschaften der Objekte jedoch besonders berücksichtigt. Ist zum Beispiel die Farbe einer der Eigenschaften, die repräsentiert werden soll, dann wird bei dieser Art von Isomorphie die physikalische Identität des Darstellungsmediums gefordert. Hierzu stellt man sich am besten die Menge der Eigenschaften sortiert in Dimensionen vor. ‚Hat die Farbe x' bzw. ‚ist x cm hoch' sind solche Dimensionen, und es kann immer nur ein Wert aus dem Bereich einer Dimension zutreffen. Ein konkreter Isomorphismus erster Stufe sorgt nun dafür, daß die ursprüngliche Eigenschaft und ihr Repräsentant zur selben Dimension gehören. Ein abstrakter Isomorphismus unterscheidet sich nun dahingehend, daß zwar immer noch Eigenschaften auf Eigenschaften und Relationen auf Relationen abgebildet werden, aber die Gleichheit der Dimension wird durch eine solche Beziehung nicht mehr sichergestellt. Der Isomorphismus zweiter Ordnung läßt die Eigenschaften bzw. die Dimensionen überhaupt außer Betracht; er berücksichtigt nur Relationen mit einer Stelligkeit > 2." (Rehkämper 1990: 14f.)

[171] Vgl. Cooper/Shephard 1973, Metzler/Shephard 1974, Cooper 1976, Shephard/Cooper 1982; einen Überblick über das Postulat analoger Repräsentationsformen stützende empirische Befunde gibt Steiner 1988: 102ff.

[172] Weitere Belege für die empirische Wahrscheinlichkeit analoger Wissensrepräsentation im menschlichen Gehirn finden sich bei Goschke/Koppelberg 1990.

[173] Argumente, die diese Annahme eher unwahrscheinlich machen, finden sich bei Helm 1991, Abschnitt 6.3.

Modellierung verstehen. Für andere Eigenschaften wird in der theoretischen Erklärung auf propositionale Formate zurückgegriffen werden müssen.

Ein Vertreter der schwachen These ist Johnson-Laird. Er versucht allerdings auch für klassische Beispiele der mentalen Operation mit propositionalem Gehalt *(„reasoning")* empirische Belege zu liefern, die eine Subsumtion sogar solcher kognitiven Leistungen unter „mentale Modellierung" nahelegen.[174] Ich möchte an dieser Stelle jedoch nicht weiter auf die Debatte zwischen Vertretern verschiedener Theorien über mentale Repräsentationsmechanismen eingehen, sondern mich gleich dem nächsten wesentlichen Charakteristikum mentaler Modelle zuwenden.

(3) Mentale Modelle sind simulationsfähig.

Eine wichtige Anforderung an die phänomenales Bewußtsein unterstützenden Repräsentate ist, daß sie Elemente mentaler *Simulationen* werden können. Mentale Modelle müssen interne Strukturen erzeugen können, die ganz oder teilweise mögliche Welten darstellen. Instrumente zur Lösung solcher epistemischen Aufgaben können nur Repräsentate sein, die auch unabhängig vom Strom des Inputs aktivierbar sind. Diese Tatsache läßt möglicherweise Schlüsse auf die Natur ihrer physischen Realisierung zu; diese Schlüsse sind jedoch nicht von philosophischem Interesse. Von philosophischem Interesse ist allerdings das Faktum, daß mentale Modelle durch ihre relationale Vernetzung untereinander zur Aktivierung komplexerer Modelle beitragen können. Auf diese Weise können auch kausale Relationen zwischen externen Repräsentanda intern simuliert und in zeitlichen Sequenzen „durchgespielt" werden: Wird ein gewisses Simulat (zum Beispiel eines aktuell nicht gegebenen Zielzustandes) aktiviert, so ändert sich auch der Gehalt *anderer* über ihre relationale Struktur mit ihm verknüpfter mentaler Modelle. Durch Meta-Repräsentation können dann die so entstandenen Makro-Modelle auf interessante Aspekte oder Eigenschaften hin überprüft werden.

Mentale Modelle geben uns auf diese Weise eine „Kompetenz für Kontrafaktizität", weil sie uns als psychologische Subjekte aus der Bindung an das rein sinnlich gegebene „Jetzt" lösen. Wenn wir in das Gesicht einer Freundin oder eines Freundes schauen, können wir gleichzeitig mit der sinnlichen Wahrnehmung ihres mimischen Ausdrucks frühere ihrer Gesichtsausdrücke innerlich assoziieren und Revue passieren lassen. Dabei werden uns häufig frühere Situationen einfallen, in denen sie (oder auch eine andere Person) einen ähnlichen Gesichtsausdruck hatten.[175] Solche

[174] Vgl. Johnson-Laird 1983.

[175] Konnektionistische Systeme führen eine inhaltsbezogene Speicherung von Information durch und erzeugen so ein kontextsensitives „assoziatives Gedächtnis". Ähnliche Repräsentate (Aktivierungsvektoren) werden in einem gegebenen physikalischen konnektionistischen System auch ähnliche *kausale Rollen* und ähnliche physikalische *Aktivierungsbedingungen* innehaben – dies ist die Erklärung für das oben angedeutete psychische Phänomen von „assoziierten mentalen Simulationen". Vgl. hierzu Helm 1990: 184ff.

durch aktuelle Ereignisse ausgelösten phänomenalen Simulationen können unter Umständen sehr dominant werden und die aus aktuellem Input erzeugten mentalen Modelle überlagern. Ihnen fehlt jedoch in den meisten Fällen eine wichtige Eigenschaft: der *Präsentationsaspekt*. Diese fehlende Eigenschaft verhindert, daß Simulate zu Halluzinaten werden.

(4) Mentale Modelle können einen puren Präsentationsaspekt besitzen.

Mentale Modelle können qualitativen Gehalt instantiieren. Das tun sie genau dann, wenn sie als Datenstrukturen nicht nur die relationale Struktur ihres Repräsentandums wiedergeben, sondern auch das besitzen, was ich oben als den „Signalaspekt" beschrieben habe. In Standardsituationen ist das, was das mit geschlossenen Augen imaginierte oder mit offenen Augen assoziierte mentale Modell eines vergangenen Gesichtsausdrucks meiner Freundin oder meines Freundes von dem aktuellen, direkt durch retinalen Input verursachten mentalen Modell des Gesichts meiner Freundin oder meines Freundes unterscheidet, eben genau dies: Der prägnante *qualitative* Aspekt. Wenn Sie kein Eidetiker oder Visualisierungsexperte sind, dann wird das imaginierte Modell ihres Gegenübers nämlich nicht farbig sein. Das kommt daher, daß es nicht direkt durch den aktuellen Informationsfluß von der Netzhaut aktiviert wird. Mentale Modelle können durch eine Vielzahl interner Vorgänge aufgerufen werden, zum Beispiel durch höherstufige kognitive Funktionen wie Gedächtnis und Erinnerung. Im allgemeinen zeigt die Tatsache, daß ein mentales Modell neben dem strukturell-relationalen Aspekt noch einen puren Präsentationsaspekt besitzt – die Farbigkeit, eine Klangqualität, emotionale Tönung, Schmerzhaftigkeit usw. – zwei wichtige Zusammenhänge auf:

- es wird primär aus einem bestimmten physischen Subsystem heraus aktiviert (zum Beispiel aus dem visuellen oder auditorischen Cortex, aus dem limbischen System oder den Schmerzzentren); und
- sein Repräsentandum ist *aktuell* gegeben, es modelliert Teile der *wirklichen* Welt.

Nun gibt es eine ganze Reihe von psychischen Ausnahmesituationen, in denen mentale Modelle konstruiert werden, die ganz offensichtlich nur Simulate sind und doch den fraglichen Signalaspekt besitzen. Paradebeispiele hierfür sind Träume, Visionen, Halluzinationen, hypnagoge Bilder, Phosphene[176] usw. All dies sind jedoch Fälle, in denen man davon ausgehen kann, daß die verantwortlichen Signale intern erzeugt und vom System

[176] Phosphene sind Farbflecken im Sehfeld, die nicht dadurch entstehen, daß die Netzhaut gereizt wird. Sie können dadurch ausgelöst werden, daß man bei geschlossenen Augen mit den Handflächen auf die Augäpfel drückt und so die Blutzufuhr unterbricht – aber auch durch direkte elektrische Stimulation des visuellen Cortex oder, wie man aus den Berichten von Astronauten weiß, durch kosmische Strahlung. Vgl. Hardin 1988: 94; Hurvich 1981: 14. Obwohl Phosphene eindeutig durch interne Signalquellen ausgelöst werden, erscheinen sie subjektiv als vor den Augen lokalisiert.

nicht als „Eigenreize“ erkannt werden. Träume entstehen zum Beispiel wahrscheinlich dadurch, daß Bereiche des Hirnstamms enthemmt werden und ein mehr oder weniger chaotisches Signalgewitter auszusenden beginnen, das dann durch höhere Bereiche zu einer mehr oder weniger konsistenten Geschichte interpretiert wird.

Da der Präsentationsaspekt keine Relationen zwischen Repräsentanda abbildet, geht er auch nicht in eine funktionale Repräsentation ein – er gehört also auch nicht notwendigerweise zu der relationalen Struktur, die ein mentales Modell wesentlich ausmacht. *Reine* Präsentate sind überhaupt keine mentalen Modelle, da sie weder multimodal, noch relational strukturiert oder simulationsfähig sind. Die Farbflecken, die wir kurz vor dem Einschlafen bei geschlossenen Augen sehen können, oder plötzlich und unvermittelt auftretende Stimmungen sind keine mentalen Modelle. Sie signalisieren die Präsenz einer gewissen Reizquelle, aber ohne sie relational mit mentalen Repräsentaten zu vernetzen, die durch andere Reizquellen aktiviert werden. Häufig besitzen solche Zustände keine erkennbare Funktion *für* das System. Deswegen kann man einige von ihnen auch als Artefakte oder höherstufiges „Rauschen“, d. h. als ungeordnete Aktivität im System betrachten.

(5) Mentale Modelle sind ineinander einbettbar.

Mentale Modelle sind komplexe Datenstrukturen, die ineinander eingebunden werden können. Sie können – innerhalb einer Modalität, aber auch multimodal – zu reicheren und umfassenderen mentalen Strukturen verbunden werden. Das Modell des Buches ist in das Modell der es haltenden Hände eingebunden; die modellierten Hände sind bruchlos mit dem multimodalen Repräsentat ihres[177] Körpers verbunden. Der Körper ist – falls Sie sich gerade in einem Haus befinden sollten – in einem komplexen mentalen Modell des Zimmers, in dem Sie sich jetzt aufhalten, phänomenal lokalisiert. Auf diese Weise können mentale Modelle zu Bausteinen von Makro-Modellen werden und in immer höherstufigere und reichere relationale Datenstrukturen eingebunden werden. In nicht-pathologischen Standardsituationen gibt es ein oberstes mentales Modell, das alle anderen enthält. Dieses werde ich ab jetzt als das jeweilige *Weltmodell* oder *Realitätsmodell* des Systems bezeichnen. In Abschnitt 2.3 werde ich auf die Frage zurückkommen, was es bedeutet, daß Gehirne mentale Modelle der Wirklichkeit erzeugen.

(6) Mentale Modelle können Objekte von Meta-Repräsentation werden.

Hierbei handelt es sich um einen zweiten Sonderfall der eben angesprochenen wechselseitigen Einbettbarkeit. Der **erste** Fall war mit dem Stichwort „Multimodalität“ bereits angedeutet: Mentale Modelle können Strukturen unterschiedlicher Formate repräsentational integrieren. Der **zweite** Fall er-

[177] Warum es *ihr* Körper ist, wird im nächsten Kapitel deutlich werden.

gibt sich aus der Tatsache, daß mentale Modelle Inhalte von Bewußtsein sein können.

Wenn es richtig ist, daß man über den Begriff der „Meta-Repräsentation" eine erfolgversprechende naturalistische Perspektive auf den „Bewußtheits"-Aspekt der Subjektivitätsfrage einnehmen kann, dann muß es einen spezifischen neuronalen Mechanismus der Meta-Repräsentation geben, der im Falle menschlichen Bewußtseins dafür verantwortlich ist, daß manche mentalen Modelle ganz oder teilweise zu *bewußten* mentalen Modellen werden. Die Suche nach einem solchen Mechanismus, dessen Aktivierung als hinreichende und notwendige empirische Bedingung für die Entstehung der fraglichen psychischen Qualität gelten kann, ist nicht Aufgabe einer Philosophie des Geistes. Im gegenwärtigen Stadium der Erkenntnisse über unsere eigenen inneren Zustände besteht die Rolle der Philosophie primär in einer präzisierenden Analyse der Explananda. Wenn das Explanandum die Instantiierung einer durchgängigen psychischen Qualität („Bewußtheit") relativ zu einer Klasse von mentalen Modellen und wenn die explanatorische Strategie „Meta-Modellierung" ist, dann kann man die provisorische Hypothese aufstellen, daß „Bewußtheit" durch genau all jene Datenstrukturen[178] instantiiert wird, die in diejenige Partition eines obersten und umfassendsten Modells der Welt eingebettet sind, die durch einen für jeden Systemtyp gesondert zu beschreibenden physischen Mechanismus der Meta-Repräsentation erfaßt werden. Diese Funktion bzw. der sie realisierende Mechanismus selbst ist niemals bewußt, er *macht* die in den entsprechenden Teil des Realitätsmodells eingebundenen Modelle zu bewußten mentalen Modellen.

(7) Mentale Modelle sind immer interpretiert.

Mentale Modelle zeichnen sich durch eine Eigenschaft aus, die der bereits erwähnten van Gulickschen semantischen Transparenz eng verwandt sind. Subjektiv ist ihr Gehalt bereits fixiert, wir wissen „immer schon" was der Gehalt der unser phänomenales Bewußtsein ausmachenden mentalen Modelle ist. Das mentale Modell des Buches ist unter Standardbedingungen in seinem Gehalt genauso festgelegt wie das mentale Modell ihrer es jetzt gerade haltenden Hände. Beide Instantiierungen von mentalem Gehalt sind frei von jeder Ambiguität und zwar in einem so hohen Ausmaß, daß Sie – um noch einmal die van Gulicksche Metapher zu bemühen – sie nicht *als* mentale Modelle erkennen, sondern geradewegs *durch sie hindurchschauen*. Man kann diese Einsicht auch so ausdrücken: Im Gegensatz zu satzartigen Strukturen enthalten mentale Modelle niemals Variablen.[179]

[178] Selbst wenn eine solche Instantiierungserklärung einmal gelingen würde, bestünde immer noch das Problem, daß die durch sie spezifizierten und präzisierten metarepräsentationalen Eigenschaften subpersonale Eigenschaften des Systems (dem logischen Subjekt von Instantiierungserklärungen) darstellen.

[179] Vgl. McGinn 1989, Rehkämper 1990.

Selbstverständlich gibt es nicht-interpretierte mentale Strukturen. Dazu gehören zum Beispiel visuelle Perzepte („unverstandene" Elemente aus den frühen Stadien visueller Wahrnehmung), abstrakt-geometrische Halluzinationen[180] – wie sie manchmal kurz vor dem Einschlafen oder nach Einnahme psychoaktiver Substanzen erlebt werden können – oder auch durch Alkoholkonsum oder Boxhiebe verursachte phänomenale Sternenhagel. Besonders interessant scheinen in diesem Zusammenhang ambige mentale Strukturen, die durch „unterdeterminierte Datensätze" wie zum Beispiel das Bild eines Neckerschen Würfels hervorgerufen werden. Interessant am Beispiel des Neckerschen Würfels ist zweierlei. **Erstens** können wir ihn zu einem gegeben Zeitpunkt *t* immer nur in einer Interpretation erleben. **Zweitens** scheint es uns unmöglich, mehrdeutige Bilder wie etwa den Neckerschen Würfel absichtlich als mentales Bild zu *generieren*.

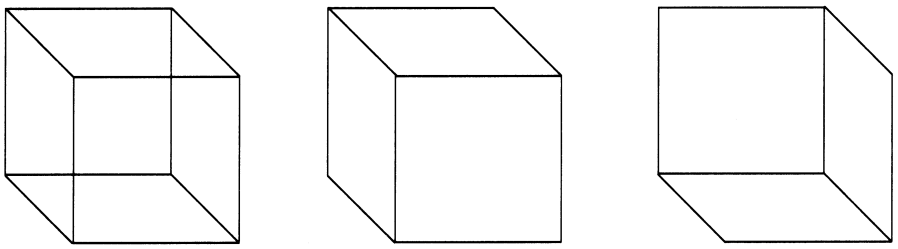

Abb. 1: Die 1832 von dem Schweizer Naturforscher L. A. Necker erfundenen „Neckerschen Würfel". Die zweidimensionale Darstellung des Würfels im ersten Bild gibt dem menschlichen Sehsystem die Möglichkeit, zwischen zwei dreidimensionalen Interpretationen (dargestellt im zweiten und dritten Bild) hin- und herzuspringen.

Das bedeutet, daß menschliche Gehirne zu einem gegebenen Zeitpunkt offensichtlich aus einem Perzept oder einem bestimmten Input immer nur *ein* mentales Modell konstruieren können, auch wenn die einströmende Information die Konstruktion eines oder mehrerer alternativer Modelle bzw. Interpretationen erlauben würde. Haben Sie schon einmal längere Zeit das Bild eines Neckerschen Würfels betrachtet? Wenn Sie es sorgfältig tun, können Sie beobachten, wie zwei mentale Modelle sich abwechseln. Sie können aber manchmal auch erleben, daß es zwischen der Aktivierung des ersten mentalen Modells und der des zweiten, alternativen Modells einen kurzen Zeitraum gibt, in denen ein nicht-interpretiertes Perzept ihr phänomenales Bewußtsein in diesem Bereich ausfüllt.[181]

[180] Vgl. hierzu Siegel 1975.

[181] Die Dauer dieses Zeitraums hängt für ein gegebenes System (wie etwa den visuellen Cortex) von der Länge des Weges ab, die das System durch seinen Zustandsraum von einem stabilen Zustand in den anderen zurücklegen muß. Die Geschwindigkeit, mit der es von einer Interpretation zur nächsten wechseln kann, hängt von seinem globalen energetischen Zustand (seiner inneren „Landschaft") und damit letztlich von *physikalischen* Bedingungen ab.

Sie werden außerdem bemerken, daß die dem Wechsel vom Modell zum Perzept zum Modell zugrundeliegenden Vorgänge neuronaler Informationsverarbeitung ihnen nicht wiederum in Form eines bewußten mentalen Modells gegeben sind. Aus diesem Grund sind sie auch nur sehr schwer durch bewußt eingeleitete mentale Operationen zu beeinflussen. Insbesondere werden Sie **zweitens** feststellen können, daß Sie sich einen ambigen Neckerschen Würfel nicht absichtlich vorstellen können. Das inputunabhängig erzeugte mentale *Simulat* des Würfels ist also immer interpretiert.[182] Dies spricht dafür, daß mentale Modelle interpretiert abgespeichert werden bzw. aus interpretierten Elementen aufgebaut werden. Es zeigt aber auch eine wichtige Gemeinsamkeit zwischen mentalen Modellen und mentalen Simulaten. Ihre Interpretation mag wechseln, aber sie ist zu einem gegeben Zeitpunkt immer determiniert. Zumindest gilt dies für jene Teilmenge von ihnen, die metarepräsentational erfaßt und dadurch zu Bewußtseinsinhalten werden.

(8) Mentale Modelle sind keine phänomenalen Individuen.

Mentale Modelle sind hypothetische Entitäten, um gewisse psychologische Eigenschaften genauer zu beschreiben und möglicherweise einer Naturalisierung – das soll heissen: einer Behandlung als empirisches Problem – zugänglich zu machen. Vielleicht sind mentale Modelle auf begrifflicher Ebene zu erfassen als Aktivierungsvektoren in den inneren Schichten konnektionistischer Systeme nach dem Vorbild neuronaler Netze, die wiederum analysiert werden können als Punkte auf einer Hyperebene im Vektorraum des gesamten Zustandsraums des Systems.[183] Möglicherweise wird die Neuroinformatik aber auch ganz andere Beschreibungen für den hier anvisierten Typ von interner Repräsentation erarbeiten. Wichtig ist dabei nur, daß der Begriff des „mentalen Modells" für zukünftige Korrekturen offen ist, die sich aus neuen Einsichten ergeben. Deswegen habe ich versucht, ihn – ohne Rekurs auf Details möglicher physischer Realisierungen - über seine repräsentationalen Charakteristika einzukreisen.

Modelle sind im Grunde *Prozesse*, nicht Objekte (und schon gar nicht metaphysische Objekte). Ihr Gehalt ist das Resultat einer permanent ablaufenden Echtzeit-Abgleichung mit dem aktuellen Input des Systems. Wenn man den phänomenalen oder den intentionalen Gehalt des stetigen Modellierungsflusses zu einem Zeitpunkt *t* betrachtet und dann in dem externen Code einer natürlichen Sprache zu beschreiben versucht, kommt es leicht zu „Reifikationen" und „Objektbildungen", die durch unser subjektives Erleben genaugenommen nicht immer gedeckt sind. Die in externen Zuschreibungen von Gehalt auftretenden Individuen (mentale Repräsentate, Simulate, Präsentate, Meta-Repräsentate, Modelle usw.) sind ontologisch neutrale Bestandteile von Erklärungsstrategien. Was an diesem Punkt häu-

[182] Vgl. Rehkämper 1990: 12, Finke 1989: 128.
[183] Vgl. Helm 1990: 177.

fig zu Konfusionen führt, ist die Tatsache, daß die Kategorie des „Objekts" eine Art von Kategorie natürlicher Zeichen darzustellen scheint, die menschliche Gehirne intern häufig verwenden, weil sie sich als funktional erfolgreich herausgestellt hat. Das sagt aber nichts über das Wesen der Wirklichkeit im allgemeinen aus, sondern nur über die repräsentationale und funktionale Architektur gewisser natürlicher Systeme. Und schließlich darf man die Tatsache, daß unser Gehirn Teile des Inputstroms *als Objekte* modelliert, nicht so mißverstehen, daß man sagt, die mentalen Modelle seien *selbst* Gegenstände – dann hätte man nicht nur eine falsche Theorie, sondern auch ein falsches mentales Modell mentaler Modelle.[184]

(9) Mentale Modelle sind funktional aktiv.

Mentale Modelle sind Strukturen, die nicht nur einen Gehalt besitzen, sondern idealerweise auch ein funktionales Profil. Von bloßen Analogrepräsentaten unterscheidet sie, daß sie eine kausale Rolle in der inneren Ökologie des sie erzeugenden Systems spielen. Diesen Punkt hat vor einem halben Jahrhundert bereits Kenneth Craik betont, auf den ich gleich noch einmal kurz zurückkommen werde. Da mentale Modelle Datenstrukturen sind, die in informationsverarbeitenden Systemen zum Beispiel durch eine große Anzahl gleichzeitig ablaufender (aber prinzipiell physikalisch beschreibbarer) Mikro-Ereignisse erzeugt werden, können sie unter solchen Beschreibungen auch andere physikalische Ereignisse verursachen und so einen Einfluß auf das Verhalten des Systems haben. Wenn wir eine Entität nachweisen könnten, die über ihren Gehalt individuiert wird und gleichzeitig *qua* Ereignis unter einer physikalischen Beschreibung in gesetzesartige Korrelationen eintreten kann, wenn wir also die Nomizität von mentalem Gehalt[185] demonstrieren könnten, hätten wir nebenbei das Leib-Seele-Problem gelöst.

Das Leib-Seele-Problem löst man jedoch nicht durch die Einführung programmatisch-hypothetischer Terme[186], sondern dadurch, daß diese Terme mit empirischem Gehalt gefüllt werden. Dafür ist es noch zu früh. Aus konnektionistischer Perspektive kann man jedoch sagen: Konnektionistische Systeme transformieren interne Analogrepräsentate, durch Aktivierungsvektoren beschreibbare Netzwerkzustände, in andere interne Analogrepräsentate. Betrachtet man nun das Gehirn als ein konnektionistisches System, dann wird deutlich, daß Aktivierungszustände in den motorischen Regionen andere kausale Konsequenzen haben werden als, sagen wir, Aktivierungszustände in bestimmten Bereichen des visuellen Cortex. Da manche Teile des Gehirns über efferente Nervenbahnen mit dem restlichen

[184] Man sieht, daß das die abendländische Philosophie so stark bestimmende Substanz-Attribut-Denken seine Wurzeln unter anderem in Eigenheiten natürlich entstandener, unbewußter und subpersonaler Repräsentationsfunktionen im menschlichen Gehirn besitzt. Damit ist natürlich nichts über den Status der mit ihm häufig verknüpften Ontologie gesagt.

[185] Vgl. Bieri 1990, Davidson 1981.

[186] Kriterien für eine Theorie zur Lösung des Leib-Seele-Problems habe ich in Metzinger 1985, 1990 entwickelt.

Körper und seinen Handlungsorganen verbunden sind, wird die Aktivierung mentaler Modelle („motorischer Simulate“) in diesen Bereichen andere Folgen haben als die Aktivierung mentaler Modelle in kognitiven Modulen oder Subsystemen des Gehirns. Ähnliche distribuierte Repräsentate in einem Parallelsystem werden durch ähnliche Aktivierungszustände erzeugt, die wiederum ähnliche kausale Rollen innerhalb der funktionalen Architektur des Systems spielen. Das Aufregende an dem Begriff der „distribuierten Repräsentation“ ist nicht nur, daß er wahrscheinlich in einem gewissen Ausmaß neurobiologisch realistisch ist[187] und daß man den Begriff des „mentalen Modells“ allem Anschein nach unter ihn subsumieren kann. Wirklich brisant wird er dadurch, daß er uns ein mathematisch exaktes Ähnlichkeitsmaß anbietet, nachdem wir nicht nur die *Ähnlichkeit des Gehalts* zweier Repräsentate beurteilen können, sondern – falls wir die physikalischen Eigenschaften des sie aktivierenden physikalischen Systems gut genug kennen – auch eines für die *Ähnlichkeit der kausalen* Rolle.[188] Mit anderen Worten: Ähnliche mentale Simulationen in den motorischen Regionen werden – *ceteris paribus* – auch zu ähnlichem motorischem Output führen. Handeln besteht in der Aktivierung und Manipulation einer ganz bestimmten Klasse von mentalen Modellen, nämlich solchen mentalen Modellen, die über ihre neurobiologischen Aktivierungsbedingungen mit den physikalischen Ursachen gewisser Verhaltensmuster gekoppelt sind.

Fassen wir kurz zusammen. Mentale Modelle sind multimodale Analog-

---

[187] Vgl. jedoch Fußnote 2. Mentale Modelle vereinigen im Gegensatz zu internen Satzanaloga Repräsentations- und Steuerfunktion, weil sie einen Teil der relationalen Struktur ihres Objektes simulieren. Sowohl unsere neuronale *wetware* als auch mentale Modelle weisen eine komplexe relationale Struktur auf: Erregungsmuster in Gehirnen kann man als dynamische relationale Datenstrukturen in Biosystemen begreifen, die ihren intentionalen Gehalt über ihre relationale Struktur darstellen und umwandeln. Das Verhalten des Systems wird dadurch gesteuert, daß diese Veränderungen gleichzeitig den Informationsfluß und auch komplexe physikalische Eigenschaften des Systems verändern.

[188] Über die für eine Modellierung kognitiver Fähigkeiten interessanten Eigenschaften distribuierter Repräsentate schreibt Gerhard Helm: *„Alle diese Eigenschaften beruhen im Grunde genommen auf einem Punkt: Die Repräsentationen können als Vektoren betrachtet werden, für die es . . . ein mathematisch exaktes Ähnlickeitsmaß gibt. [FUSSNOTE: Neben dem oben erwähnten Winkel zwischen zwei Vektoren im Vektorraum wird oft auch die Distanz zwischen zwei Punkten betrachtet, die die (normierten) Vektoren im Vektorraum repräsentieren. Beide Ähnlichkeitsmaße sind äquivalent, d. h., sie führen nicht zu unterschiedlichen Ähnlichkeitsbeziehungen zwischen verschiedenen Vektoren.] Daraus ergibt sich, daß die intuitive semantische Ähnlichkeit von Begriffen, die sich in der Anzahl der gemeinsamen „features“ niederschlägt, mathematisch exakt dargestellt (und ausgewertet) werden kann. Die semantische Ähnlichkeit zweier Begriffe spiegelt sich daher direkt in der repräsentierenden Struktur (in der „Syntax“) wider. Betrachtet man einen Vektor einmal als eine Art von „Symbol“ (in dem Sinn, daß ein Symbol etwas ist, was eine Bedeutung trägt;. . .) so ist leicht einzusehen, wie bei konnektionistischen Systemen „die kausale Rolle mit dem Inhalt zur Deckung gebracht wird, indem man die Parallelen zwischen der Syntax eines Symbols und seiner Semantik ausnutzt.“ [FUSSNOTE: Fodor (1987), S. 19 (Fodor selbst bezieht sich mit dieser Äußerung natürlich auf klassische Architekturen).] Semantisch ähnliche Begriffe werden durch ähnliche Vektoren repräsentiert und ähnliche Vektoren spielen auch eine ähnliche kausale Rolle im System.“* Vgl. Helm 1990: 184.

strukturen, die von informationsverarbeitenden Systemen (wie etwa biologischen Gehirnen) erzeugt werden. Sie bilden Teile der relationalen Struktur ihrer Originale über gewisse höherstufige Isomorphismen ab. Mentale Modelle können vom System zu internen Simulationsläufen eingesetzt werden und sind wechselseitig ineinander einbettbar. Durch diese Eigenschaften können sie zu Bausteinen komplexer mentaler Strukturen werden. Sie können, in Nicht-Standardsituationen oder wenn ihre Repräsentanda aktuell als kausale Auslöser gegeben sind, einen qualitativen oder Präsentationsaspekt besitzen und sie können durch Meta-Repräsentation zu Bewußtseinsinhalten werden. Mentale Modelle sind immer interpretiert und werden wahrscheinlich interpretiert abgespeichert[189], sind aber nicht notwendigerweise auf eine Interpretation festgelegt. Individuiert werden sie durch den Gehalt eines spezifischen inneren Repräsentationsprozesses zu einem gegebenen Zeitpunkt.

Ich meine, mentale Modelle sind aus einer Reihe von Gründen interessant für eine Theorie des Geistes. Der hier knapp skizzierte Begriff des mentalen Modells öffnet eine neue Beschreibungsebene für informationsverarbeitende Systeme wie uns selbst. Wir können Systeme als *Modellgeneratoren* beschreiben, als Maschinen, die Simulationen der Welt – und ihrer selbst – erzeugen. Psychische Phänomene lassen sich nun als mentale Modellierungen analysieren und möglicherweise eröffnet uns diese neue Sichtweise auf den psychophysischen Problemkomplex die Formulierung präziserer Fragen. Mentale Modelle sind nämlich eine spezielle Klasse interner Repräsentate. Die Beschreibungsebene, die wir durch diesen besonderen Typ interner Strukturen gewinnen, liegt deutlich unter der personalen Beschreibungsebene: Mentale Modelle sind nicht das, was wir traditionell unter Personen verstehen und die Eigenschaften mentaler Modelle sind nicht die Eigenschaften von Personen.[190]

Andererseits liegt die Ebene mentaler Modellierung klar über der Ebene funktionaler Zustände, weil mentale Modelle durch ihren Gehalt individuiert werden, funktionale Zustände dagegen über ihre kausale Rolle. In einer funktionalistischen *top-down*-Analyse informationsverarbeitender Systeme stellt man die Frage: *Welche neurobiologischen oder physischen Zustände des Systems haben dieselbe kausale Rolle inne wie der relevante funktionale Zustand?* In bezug auf mentale Modelle kann man diese Frage nur noch dann sinnvollerweise aufwerfen, wenn man starke Zusatzannahmen über die Architektur des Systems macht. Im allgemeinen muß man im Sinne einer Instantiierungserklärung fragen, welche zugrundeliegenden Strukturen unbedingt erforderlich sind, um das jeweilige Modell zu aktivieren. Mentale Modelle sind komplexe Datenstrukturen eines bestimmten Typs, die von manchen Systemen intern erzeugt werden. Sie sind außerdem Entitäten einer naturalistischen Theorie des Geistes, weil man davon aus-

[189] Das könnte zum Beispiel durch die strukturelle Einbettung in die innere Gesamtkonfiguration des Systems erfolgen.

[190] Man kann jedoch die von Personen intern erzeugten mentalen Modelle ihrer selbst als „phänomenale Personen“ beschreiben. Vgl. Abschnitt 3.2.1.

geht, daß sie durch natürlich entstandene Mechanismen generiert werden und in Form physischer Strukturen realisiert sind. Mit ihrer Realisierung geht die Instantiierung bestimmter psychologischer Eigenschaften einher. Und doch hat es den Anschein, als ob „Realisierung" für mentale Modelle etwas anderes heißt als für funktionale Zustände, so wie sie seit den frühen Arbeiten von Putnam[191] in der analytischen Philosophie des Geistes diskutiert worden sind. Das liegt daran, daß möglicherweise die physischen Aktivierungsbedingungen mancher mentaler Modelle (und vor allem mentaler Präsentate) so hochspezifisch sind, daß ein bestimmtes mentales Modell auch nur auf einer einzigen physischen Struktur im Universum erzeugt werden kann. In einem solchen Fall wären mentale Modelle – im Gegensatz zu funktionalen Systemzuständen – auch nicht *multirealisierbar.*[192]

Der Begriff des „mentalen Modells" ist das Herzstück einer Theorie geistiger Zustände, die man als die „Cambridge-Theorie der mentalen Repräsentation"[193] bezeichnen könnte. Als geistiger Vater dieser Theorie darf Kenneth Craik gelten. Er veröffentlichte im Jahre 1943 ein Buch unter dem Titel *„The Nature of Explanation"*, das in seinen Thesen quer zur damaligen behavioristischen Euphorie lag, weil es zur Erklärung kognitiver und behavioraler Leistungen starke Annahmen bezüglich interner Strukturen machte. Craik nahm bereits vor dem Aufstieg des Computers als einer technisch-theoretischen Metapher an, daß Menschen Umweltereignisse in interne Strukturen überführen, an diesen Strukturen gewisse Manipulationen vornehmen und diese dann wieder in Handlungen transformieren.

Charakteristisch für den frühen Begriff des Modells, wie er von Craik entwickelt wurde, ist genau jene Ähnlichkeit der relationalen Struktur[194],

---

[191] Vgl. Putnam 1975; aber auch seine heutige Kritik am Funktionalismus als einer Theorie des Geistes in Putnam 1991, Kapitel 5 und 6, sowie 204ff.

[192] Vielleicht kann man einen Ausweg aus diesem Dilemma finden, indem man den Übergang von einem Maschinenfunktionalismus im Sinne von Turing und Putnam zu einem konnektionistischen *Mikrofunktionalismus* vollzieht und dann die folgende These aufstellt: Mentale Modelle werden durch Mengen mikrofunktionaler Zustände realisiert, von denen jeder einzelne mit einem konkreten neurobiologischen Mikrozustand identisch ist. Wenn wir über den Gehalt eines mentalen Modells sprechen, sprechen wir also über eine Menge von mikrofunktionalen Zuständen auf dem subsymbolischen Niveau eines Systems. Diese Mengen sind aber so komplex, daß sie möglicherweise auch innerhalb der Lebensgeschichte eines einzigen Systems nur ein einziges Mal instantiiert werden. Worauf wir bei der Zuschreibung von mentalem Gehalt Bezug nehmen, sind deshalb Schnittmengen aus einer großen Anzahl von Mengen mikrofunktionaler Zustände, prototypische Idealisierungen eines Ensembles von neuronalen oder physikalischen Mikroereignissen mit einem besonders prägnanten funktionalen Profil.

[193] Diesen Ausdruck habe ich von Colin McGinn übernommen. Vgl. McGinn 1989: 178.

[194] Bezüglich der Schwierigkeiten traditioneller Versuche, den philosophischen Begriff der „Ähnlichkeit" näher zu bestimmen vgl. Scholz 1991: 43ff. Scholz weist auf eine Schwierigkeit für isomorphistische Repräsentationstheorien hin: *„Strukturähnlichkeit ist – genauso wie Ähnlichkeit – eine reflexive und symmetrische Relation. (Zusätzlich ist Strukturähnlichkeit transitiv.) Da dies auf die Darstellungsbeziehung nicht zutrifft, kann sie nicht einfach in der Isomorphiebeziehung bestehen;. . ."* (Scholz 1991: 58) Isomorphismen als alleiniges logisch-strukturelles Fundament einer echten epistemischen Beziehung sind zu schwach: Sie benötigen die Einbettung in einen kausal-teleologischen Entstehungs- und Anwendungzusammen-

die ich oben als Merkmal analoger Repräsentate gekennzeichnet habe. Dank eines solchen Isomorphismus zweiter Ordnung funktionieren interne Modelle auf eine spezifische Weise, die sie zu simulationsfähigen Repräsentaten macht. Ein heutiger philosophischer Vertreter der „Cambridge-Theorie", Colin McGinn, schreibt über die Grundlegung des Begriffs eines mentalen Modells durch Kenneth Craik:

> What Craik adds to the bare idea of analogue representation is the insistence that mental models function in a certain way, that they are working models that permit simulation experiments to be performed on them. But still these models are to be conceived analogically not digitally; in fact; they must be so conceived if they are to count as genuine *models*, since digital representations (. . .) do *not* have the same ‚relation-structure' as what they signify. Internal sentences would indeed be digital representations, but models are not sententially structured (. . .).[195]

Prominentester Vertreter einer Theorie mentaler Modellierung ist heute der ursprünglich ebenfalls in Cambridge ansässige Kognitionsforscher Philip Johnson-Laird. Er hat den Begriff des mentalen Modells weiterentwickelt[196] und unterscheidet eine Reihe von Anwendungen in verschiedenen Bereichen der Theoriebildung. So kann man zum Beispiel mentale Modelle der *Wahrnehmung* betrachten. Solche Modelle stellen überwiegend Information aus den Sinnesorganen intern dar, integrieren Repräsentate aus verschiedenen Sinnesmodulen und allgemeines Hintergrundwissen über die Struktur der Welt zu einer höherstufigen mentalen Struktur. Für diesen Typ von mentalem Modell sind besonders die räumlichen Beziehungen zwischen einzelnen Wahrnehmungsobjekten von Bedeutung: Für das visuelle Modell der Welt etwa wird das zweidimensionale Erregungsmuster auf der Netzhaut durch komplexe Operationen in ein dreidimensionales Modell der das Licht reflektierenden äußeren Gegenstände transformiert.[197] Dieses Modell gibt die räumlichen Beziehungen zwischen den einzelnen optischen Reizquellen und ihre Oberflächen- bzw. Reflektionseigenschaften wieder. Nach der „Cambridge-Theorie" der mentalen Repräsentation ist dieses Modell direkt verantwortlich für den Gehalt unseres phänomenalen Bewußtseins und als Erlebnissubjekte können wir es niemals transzendieren. Johnson-Laird formuliert diesen Grundgedanken bezüglich sensorisch dominierter mentaler Modelle folgendermaßen:

---

hang. In der Einführung des Begriffs der mentalen Repräsentation zu Beginn dieses Kapitels spielt das teleologische Zusatzkriterium eben diese Rolle der Abschirmung der Isomorphietheorie gegen „Trivialisierungsargumente". *„Die Schwierigkeit ist mithin nicht, daß Bild und Abgebildetes nicht isomorph sind, sondern daß sie dieses Merkmal noch nicht von anderen Komplexen unterscheidet. Der rein formale oder logische Begriff der Isomorphie muß durch empirische Auflagen verstärkt werden, wenn er Bild-Abgebildetes-Paare von anderen unterscheiden soll."* (Scholz 1991: 60)

[195] Vgl. McGinn 1989: 179.

[196] Vgl. zum Beispiel Johnson-Laird 1983 und 1988.

[197] Vgl. dazu auch Eimer 1990, Hardin 1988, Marr 1982.

> Our phenomenological experience of the world is a triumph of natural selection. We seem to perceive the world directly, not a representation of it. Yet this phenomenology is illusory: what we perceive depends on both what is in the world and what is in our heads – on what evolution has „wired" into our nervous systems and what we know as a result of experience. The limits of our models are the limits of our world.[198]

Außerdem spricht Johnson-Laird von sogenannten mentalen „Diskurs-Modellen"[199], die propositionalen Strukturen zugrundeliegen und uns ermöglichen, die beiden unterschiedlichen Formen mentalen Gehalts miteinander zu verknüpfen. Diskurs-Modelle explizieren nicht die Struktur von Sätzen, sondern die von Situationen in unserer Wahrnehmung oder Vorstellung. Johnson-Laird bietet eine Reihe empirischer Belege für die parallele Existenz von linguistischen Repräsentaten und Diskurs-Modellen an, welche ich hier nicht diskutieren werde.[200] Wichtig ist jedoch der resultierende „psychologistische Wahrheitsbegriff": Ein Diskurs bzw. die ihm entsprechenden Behauptungen werden vom System genau dann als wahr beurteilt, wenn ihr mentales Modell in das gegenwärtige Modell der Welt *eingebettet* werden kann. „Einbettung" bedeutet hier eine repräsentationale Integration, bei der Individuen mit all ihren Relationen und Eigenschaften erhalten bleiben. Diskurse referieren also auf die mentalen Modelle der von ihnen beschriebenen Situationen. Die ursprüngliche linguistische Repräsentation trägt auf diese Weise zusammen mit den jeweiligen Konstruktionsmechanismen für Diskurs-Modelle die Bedeutung des Diskurses – die Menge aller möglichen von ihr beschreibbaren Situationen. Als *wahr* beurteilt wird ein Diskurs immer dann, wenn mindestens eines seiner mentalen Modelle erfolgreich in das Modell der Welt eingebettet werden kann.[201]

Johnson-Laird geht ebenfalls davon aus, daß propositionale Einstellungen durch mentale Modelle dargestellt werden können, daß sie eine wichtige Rolle bei syllogistischen Schlüssen und in der Aufrechterhaltung komplexer Wissensstrukturen[202] spielen. Aus all dem wird deutlich, daß der

---

198 Vgl. Johnson-Laird 1989: 470f.
199 Vgl. Johnson-Laird 1989: 471.
200 Vgl. Johnson-Laird 1983.
201 Vgl. Johnson-Laird 1989: 475.
202 Es gibt eine ganze Reihe von Hypothesen zur Repräsentationsform komplexer Repräsentanda und Wissensbereiche. Vorgeschlagen werden zum Beispiel „Frames" (Minsky 1975), „Scripts" (Schank/Abelson 1977) und „Schemas" (Rumelhart 1980). Es handelt sich hierbei um unbewußte mentale Strukturen, die mit dem in das System eintretenden episodischen Informationsfluß in Wechselwirkung treten und durch eine Modifikation der allgemeinen im Schema enthaltenen Information eine neue Repräsentation erzeugen, d.h. eine neue kognitive Struktur, ein *instantiiertes* Schema. Mit Schemata läßt sich jedoch nicht erklären, warum Menschen eine große Anzahl von komplexen Situationen bewältigen, in denen Informationen aus der Vergangenheit keine Rolle spielen. Johnson-Laird (Johnson-Laird 1983) verwendet explizit den Begriff des „mentalen Modells", der spezifische Repräsentationen meint, in denen allgemeine, zugrunde gelegte Datenstrukturen keine Rolle spielen. Mentale Modelle bilden in ihrer Struktur die relevanten Eigenschaften des jeweiligen Weltzustands ab. Im Gegensatz zu propositionalen Repräsentationen steht ihre Struktur in einer direkten

Begriff des mentalen Modells derzeit ein Arbeitsbegriff ist, der von Wissenschaftlern verschiedener kognitionswissenschaftlicher Disziplinen in unterschiedlichen Problembereichen eingesetzt wird. Er ist nicht vollständig präzisiert und kann es beim derzeitigen Stand unseres Wissens über mentale Zustände auch gar nicht sein. In der Version von Johnson-Laird, einem der wichtigsten zeitgenössischen Proponenten, bildet er zusammen mit linguistischen Repräsentaten und Verarbeitungsmechanismen für beide Formattypen das zentrale Element einer empirisch orientierten Theorie mentaler Repräsentation. Die drei wichtigsten Charakteristika eines mentalen Modells bei und in den Worten von Johnson-Laird sind:

> 1. Its structure corresponds to the structure of the situation it represents.
>
> 2. It can consist of elements corresponding only to perceptible entities, in which case it may be realized as an image, perceptual or imaginary. Alternatively it can contain elements corresponding to abstract notions; their significance depends crucially on the procedures for manipulating models.
>
> 3. Unlike other proposed forms of representation, it does not contain variables.[203]

Ich habe in den vergangenen Abschnitten versucht, dem Begriff des „mentalen Modells" soviel Gehalt wie möglich zu geben, obwohl dies beim gegenwärtigen Stand unseres Wissens nicht vollständig möglich ist. Mentale Modelle sind empirisch plausible[204] Entitäten, die sich durch eine Reihe von Eigenschaften auszeichnen, die sie für eine naturalistische Theorie mentaler Repräsentation interessant machen. Eine dieser interessanten Eigenschaften ist die besondere Weise, in der sie ein System bei seiner Weltgerichtetheit unterstützen können. Eine zweite relevante Eigenschaft liegt in der offensichtlichen Tatsache, daß mentale Modelle auf immer höheren Stufen ineinander eingebettet werden und auf diese Weise komplette phänomenale Modelle der Welt oder Modelle der Wirklichkeit entstehen können. Diesen beiden Kennzeichen relationaler Analog-Repräsentate werde ich mich nun in den kommenden Abschnitten dieses Kapitels zuwenden, bevor ich mich dann im folgenden Kapitel wieder der Frage nach der Subjektivität mentaler Zustände widme.

Analogie oder Ähnlichkeitsrelation zum jeweiligen Repräsentandum. „Kausale mentale Modelle" sind nach Brewer (Brewer 1987) bereichsspezifische Unterklassen von Schemata und mentalen Modellen, die kausale Formen der Repräsentation im Umgang mit physikalischen Systemen benutzen. Von „Situationsmodellen" sprechen Dijk und Kintsch (Dijk/Kintsch 1983) und meinen eine Ebene der Textanalyse, die über die eigentliche Textstruktur hinausgeht. Brewer analysiert die verschiedenen Varianten des Modellbegriffs und kommt zu folgendem Vorschlag: Instantiierte Schemata sind die spezifischen Wissensstrukturen, die sich aus dem allgemeinen, in Globalschemata repräsentierten Wissen ableiten. Episodische Modelle sind dagegen spezifische Wissensstrukturen, welche neue Situationen repräsentieren und dabei mit dem in Form lokaler Schemata repräsentierten Wissen interagieren. Den Begriff des mentalen Modells selbst hält Brewer allgemein, ähnlich wie ich es hier tue.

[203] Vgl. Johnson-Laird 1989: 488.

[204] Zur Frage der empirischen Plausibilität vgl. Gentner/Stevens 1983, Johnson-Laird 1983, Schnotz 1988, Steiner 1988.

2.2.1 Digitales Wissen und analoges Wissen

Philosophen arbeiten mit der Sprache als ihrem wichtigsten Instrument und so ist es nicht verwunderlich, daß sie das Gehirn zuallererst als eine syntaktische Maschine in Anschlag bringen. Syntaktische Maschinen arbeiten mit einer internen Sprache, die aus formal spezifizierten Elementen besteht, auf denen rein syntaktisch definierte Operationen durchgeführt werden. Das bedeutet, daß innere Symbolketten ausschließlich nach Kriterien der Form manipuliert und verarbeitet werden. Wenn man sich für diesen Typ von mentalem Repräsentat als zentralem Bestandteil theoretischer Erklärungen entschieden hat, dann liegt es nahe, eine innere Sprache und interne Satzanaloga zu postulieren: Fodorsches Mentalesisch.[205] Man hat sich für einen *digitalen* Repräsentationscode entschieden und sich damit gleichzeitig einem einflußreichen Lager in der Philosophie des Geistes angeschlossen: den Sententialisten. Sententialisten sind Leute, die auf der Suche sind nach einer mechanistischen Erklärung mentalen Gehalts. Sie wollen eine Erklärung dafür anbieten, wie Intentionalität durch mechanische Operationen – zum Beispiel in menschlichen Gehirnen – realisiert sein kann. Den einzigen Kandidaten für solche Operationen sehen Sententialisten in der internen Transformation von Symbolketten.

Repräsentationssysteme, die diskrete interne Symbole mit einer Konstituentenstruktur verwenden, haben allem Anschein nach eine Reihe von Vorteilen, wenn wir höhere kognitive Leistungen wie begrifflich-propositionales Denken erklären wollen. Wenn solche Systeme nach dem Modell der Theorie von Fodor und Pylyshyn über eine *Language of Thought* mit einer kombinatorischen Syntax und Semantik verfügen, dann können wir die Produktivität und die Systematizität von Denkvorgängen, sowie die Kohärenz von Ableitungen durch Eigenschaften des von diesen Systemen intern benutzten Repräsentationscodes sehr gut erklären. Fodor und Pylyshyn haben behauptet, daß konnektionistische Systeme diese Erklärungen nicht anbieten können.[206] Wenn dies richtig wäre, dann könnten neuronale Netze viele der höheren kognitiven Leistungen, die den Menschen auszeichnen, nicht hervorbringen. Neuere Erkenntnisse scheinen eher in die entgegengesetzte Richtung zu deuten.[207] Ich werde allerdings an dieser Stel-

---

[205] Vgl. Fodor 1975, 1981, 1984, 1985, 1987, Pylyshyn 1984.

[206] Vgl. Fodor/Pylyshyn 1988. Ich werde hier nicht weiter auf die Debatte zwischen Anhängern des Konnektionismus und klassisch-kognitivistischer Strategien eingehen. Einen Überblick über Fodors und Pylyshyns Argumente, die Einwände gegen nicht regelgeleitete Systeme von Pinker und Prince (1988) und eine vorläufige Auswertung der Diskussion geben Bechtel/Abrahamsen 1991, Kapitel 7. Eine weitere Diskussion der Problematik findet sich in Helm 1991, Kapitel 5 & 6. Vgl. auch Goschke/Koppelberg 1990, (1991), Clark 1989, Kapitel 8 und 9.

[207] Einen Überblick über die der These einer *starken Kompositionalität semantischen Gehalts* widersprechenden empirischen Befunde bezüglich der Kontextsensitivität und der Instabilität von Begriffsstrukturen mit Blick auf menschliche Kognition geben Goschke/Koppelberg 1990. *„Rather than computing a complex meaning from context-free lexical entries,*

le nicht weiter auf den Glaubenskrieg zwischen den Vertretern des neuen Paradigmas und dem klassischen „von-Neumann-Kognitivismus“ eingehen, der tobt, während diese Zeilen geschrieben werden. Unser Problem ist nicht der kognitive, sondern der *phänomenale* Gehalt mentaler Zustände.

Interne Sätze haben jedoch auch dann eine Reihe von Nachteilen, wenn man sie als mögliche Bausteine für phänomenales Bewußtsein einsetzen möchte. Zum Beispiel sind sie nicht simulationsfähig. Wie sollten komplexe mentale Simulationen mit Simulaten durchgeführt werden, die ein ausschließlich propositionales Format besitzen? Außerdem kann man wesentliche Merkmale des phänomenalen Gehalts mentaler Zustände – der zu

---

people seem to use analogies and similarities between different contexts to infer the meaning of a given contextual combination.“ (Goschke/Koppelberg 1990: 261) Goschke und Koppelberg zeigen, daß man auch das kategoriale Wissen eines konnektionistischen Systems über Partitionen seines Zustandsraums analysieren kann (vgl. hierzu auch Churchland 1989). Ein über mehrere „verborgene Einheiten“ eines konnektionistischen Systems hinweglaufendes Aktivierungsmuster kann semantisch als ein *Begriff* interpretiert werden, wobei die von diesem Begriff zusammengefaßten Mikroeigenschaften keine Gegenstücke in einer natürlichen Sprache besitzen müssen. So entspricht einer bestimmten Klasse von Inputvektoren eine bestimmte Menge von Punkten im Zustandsraum des Systems, wobei man prinzipiell die Ähnlichkeitsrelationen zwischen Inputinstanzen in internen Aktivierungsvektoren beliebig scharf definieren kann. Gegen Fodor und Pylyshyn vertreten Goschke und Koppelberg eine schwache Kompositionalitätsthese, indem sie dafür argumentieren, daß ein System komplexe sprachliche Ausdrücke verstehen kann, ohne daß die vom System intern eingesetzten Repräsentationsmechanismen die Konstituentenstruktur des externen Symbolsystems wiedergeben müssen. Dieser Alternativentwurf ist empirisch recht plausibel und führt unter anderem zu der Folgerung, daß ein System, das mehrfach mit einer Instanz desselben externen symbolischen Atoms konfrontiert wird und sich auf dem Weg durch seinen Zustandsraum in das jeweils relative Optimum des energieärmsten Zustands relaxiert (also seinen Input interpretiert), dabei durchaus *unterschiedliche* stabile Endzustände erreichen kann. Einem symbolischen *token* auf der Inputebene können also diachronisch – je nach der aktuellen Globalkonfiguration des Systems – verschiedene Punkte im Zustandsraum entsprechen: Die distribuierten Repräsentationen in einem konnektionistischen System sind kontextsensitiv und die Grenze zwischen Begriff und Metapher ist eine prinzipiell fließende. Das bedeutet aber nicht, daß ein solches System auf der Performanzebene die Manipulation diskreter Symbole in einem kontextfreien kombinatorischen Medium (zum Beispiel natürlicher Sprachen) nicht perfekt simulieren könnte (Bei begrifflichem Denken könnte diese Ebene eine innere sein: die Emulation eines seriellen Prozessors *in* einem PDP-System). Die interne Simulation formaler und regelgeleiteter Operationen (Rechnen, Ableitung) müßte ein solches System durch die Übersetzung in Serien von Analogmustern (mentalen Modellen) durchführen (Vgl. Rumelhart, Smolensky, McClelland & Hinton 1986, Band 2; Goschke/Koppelberg 1990: 267, Helm 1991, Kapitel 6, Johnson-Laird 1990, Bechtel/Abrahamsen 1991) Das Verhalten eines Systems, das auf der Output-Ebene überzeugende symbolische Operationen demonstriert, interpretieren wir also immer nur approximativ als internes Befolgen von Regeln: *„Die Regelhaftigkeit im Verhalten eines Systems ergibt sich nach der Approximationstheorie nicht durch regelbefolgende Prozesse im Inneren des Systems, sondern als emergente Eigenschaft eines dynamischen Systems. . . . Einer der interessanten Punkte bei einer derartigen Betrachtungsweise ist, daß man vollständig auf die Annahme eines inneren Symbolsystems verzichten und dennoch exaktes symbolisches Verhalten erklären kann. Durch die wiederholte Anwendung vieler einfacher Mustererkennungsprozesse mit anschließender Manipulation des Ausgangsmusters sind wir (prinzipiell) in der Lage, beliebig viele Aufgaben* [vom formal-deduktiven Typ; Anmerkung TM] *zu lösen.“* (Helm 1990: 266, 275)

einem Großteil auf ihrer relationalen Struktur und ihren speziellen Formaten beruht – nicht mehr erklären. Von noch größerer Bedeutung erscheint zudem, daß mentale Modellierung einen vollkommen anderen Begriff von Intentionalität nahelegt als inneres Sprechen in einer *lingua mentis*. Diesen Punkt möchte ich auf den folgenden Seiten etwas näher untersuchen.

Die klassischen Theorien setzen Intentionalität ein, um psychische Phänomene eindeutig zu charakterisieren, von physischen Ereignissen zu unterscheiden und verschiedene Formen von mentalem Gehalt zu differenzieren. Nach Brentano sind psychische Zustände genau dadurch gekennzeichnet, daß sie ihren Gegenstand intentional in sich enthalten (im Modus der „mentalen Inexistenz") und so eine intramentale Verdoppelung der Welt leisten. In der Weiterentwicklung durch Husserl und die phänomenologische Schule wurde die Struktur von Bewußtsein dann als noetisch-noematisch analysiert: Noema ist das intentionale Objekt, der irreelle Gehalt des Bewußtseins, welcher durch temporale intentionale Akte des Vermeinens – Noesis – in eine Korrelation eintritt mit dem reellen Gehalt. Diese Korrelation zwischen Noesis und Noema ist Intentionalität. Die Akte, durch die die intentionale Bezogenheit auf nur subjektiv gegebene Objekte geleistet wird, werden von einem reinen Ich vollzogen, welches gleichzeitig die Einheit des Erlebnisraumes stiftet.[208] Es gibt also eine mysteriöse Kraft, mit der der Geist sich auf die Welt richtet und intentionale Objekte erzeugt – und diese Kraft heißt Intentionalität. Der klassische Begriff des Bewußtseins – man könnte ihn auch als die Lehre von der *direkten* Intentionalität bezeichnen – ist eine der machtvollsten Denkfiguren abendländischer Philosophie, sie bestimmt bis heute eine Großzahl von Diskussionen und viele der in ihnen explizierten theoretischen Intuitionen. Dabei kann man leicht übersehen, daß durch ihn überhaupt nichts erklärt wird.

Eine naturalistische Theorie mentaler Zustände und der durch sie mediatisierten kognitiven Leistungen hat dagegen präzise begriffliche Alternativen anzubieten, die empirisch fest verankert sind. Durch die empirische Verankerung werden naturalistische Konzepte einer Theorie des Geistes kritisierbar, angreifbar und revidierbar. Diesen substantiellen Fortschritt in Richtung auf einen Brückenschlag zwischen Natur- und Geisteswissenschaften hat uns die Kategorie der „Information" ermöglicht, weil sie das Potential einer Verbindung von Physik (Negentropie) und Erkenntnistheorie (Wissen durch Repräsentation) in sich birgt.

An die Stelle von Intentionalität tritt nun Informationsverarbeitung. Es sind physische Vorgänge, zum Beispiel in biologischen Gehirnen, die es einem System ermöglichen, mentalen Gehalt zu generieren und sich durch

[208] Bezüglich der Einheit des Erlebnisraums muß man darauf hinweisen, daß „Bewußtsein" bei Husserl dreierlei bedeuteten kann: *„. . .die empirisch vorfindliche Verwebung psychischer Erlebnisse zur Einheit des Erlebnisstroms, sodann die innere Wahrnehmung der eigenen Erlebnisse und schließlich ist „Bewußtsein" eine zusammenfassende Bezeichnung für alle psychischen Akte oder intentionalen Erlebnisse."* (Stegmüller 1978: 63)

ihn auf die Welt zu beziehen. An die Stelle von intentionalen Objekten treten intern erzeugte Datenstrukturen – etwa mentale Modelle – die eine physische Realisierung besitzen (zum Beispiel: komplexe neuronale Erregungsmuster). Für die natürliche Genese der repräsentationalen Beziehung zwischen internen Datenstrukturen und Teilen der Welt gibt es gute kausale und funktionale Erklärungen, weil diese internen Strukturen Instrumente für die jeweiligen Systeme sind. Ihre Bedeutung erhalten mentale Modelle dadurch, daß sie von einem System *benutzt* werden. Eine interessante Beobachtung ist in diesem Zusammenhang, daß Information genau in die entgegengesetzte Richtung fließt (nämlich primär in das System „hinein")[209], in die der „Pfeil der Intentionalität" zeigt. *Intendere arcum*, den Bogen des Geistes spannen und den Pfeil der Erkenntnis auf Teile der Welt richten, ist zusammen mit der Lehre von der direkten Intentionalität eine intuitiv eingängige und weitverbreitete philosophische Metapher.[210] Aber warum ist der Pfeil der Intentionalität eine viel einleuchtendere Denkfigur als das Einwärtsfliessen von Information? Dies mag daran liegen, daß unser mentales Modell der Repräsentationsbeziehung sich an dem psychischen Phänomen der Aufmerksamkeit[211] – also am subjektiven Erleben von *gerichteter* Metarepräsentation – orientiert. Mit anderen Worten: Das mentale Modell von uns selbst als kognitiven Agenten ist häufig eines von Organismen, die ihren „epistemischen Scheinwerferstrahl" nach Belieben auf die Welt und in ihr eigenes Inneres richten, von Wesen, die *als Erlebnissubjekte* die Repräsentationsbeziehung stiften.[212] Dieses Modell ist dem

[209] Das gilt nicht für interne Metarepräsentation, in der repräsentationalen „Hierarchie" des Systems ist sie eine *abwärtsgerichtete* Aktivität. Bewußtsein ist interne Intentionalität.

[210] Vgl. Dennett 1991: 333. Dennett hat auch darauf hingewiesen, daß die verschiedenen Formen des „cartesianischen Materialismus", also die Annahme einer letzten inneren Bühne, auch innerhalb repräsentationalistischer Theorien des Geistes entstehen, indem fälschlicherweise das, was er als die *intentional stance* gekennzeichnet hat, ins System hineingetragen wird; vgl. Dennett 1991: 458. Thomas Nagel hat allerdings zu Recht darauf hingewiesen, daß auch Dennetts eigene Strategie der „Heterophänomenologie" sich implizit immer noch auf die Perspektive der ersten Person verlassen muß, wenn sie nicht in einen platten Behaviorismus zurückfallen will. Vgl. Nagel 1991.

[211] Eine alternative und trotzdem verwandte Vermutung ist, daß sich das philosophische Theoretisieren über die Intentionalitätsbeziehung nach dem mentalen Bild der Welt gerichtet hat, das durch unseren stärksten Sinn erzeugt wird: Wenn mentale Repräsentation nach dem Modell der dominanten sensorischen Modalität (dem Sehen) interpretiert wird, dann gibt es automatisch – wie in unserem visuellen Realitätsmodell – *distale Objekte*. Die Gegenstände von mentaler Metarepräsentation bzw. phänomenalem Bewußtsein werden dann fast zwangsläufig eine philosophischen Deutung als systemexterne *mentale* Objekte erfahren. Für den interessanten Sonderfall der mentalen *Selbstrepräsentation* führt der Weg der verdinglichenden Analyse zu einem ganz besonderen mentalen Objekt, nämlich dem metaphysischen Subjekt – es ist gleichzeitig unendlich fern und unendlich nah.

[212] Wenn man so will ist Informationsverarbeitung „invertierte Intentionalität", weil das Gehirn die Repräsentationsbeziehung mental genau umgekehrt modelliert. Die klassischen Intentionalitätstheorien explizieren das (falsche) *mentale Modell* der Repräsentationsbeziehung. Dieses entsteht, weil die „intentionalen Akte des Vermeinens" irrtümlich dem Selbstmodell (Vgl. Kapitel 3) zugeschrieben werden und nicht dem es konstruierenden Gehirn.

tatsächlichen Sachverhalt sehr unähnlich, bestimmt aber weiterhin in sehr starkem Maß unser Denken und Selbsterleben und führt zu immer neuen philosophischen Verwirrungen, besonders was die Innerlichkeit mancher Formen von mentalem Gehalt angeht. Denn wenn unser Realitätsmodell die cartesianische Bühne beinhaltet, dann werden auch Introspektionsberichte mental wieder nach dem Modell des Berichtens über externe Ereignisse in der Welt modelliert.

Ich möchte an dieser Stelle nicht weiter auf die Überlegenheit des Informationsverarbeitungsansatzes in der Philosophie des Geistes gegenüber früheren theoretischen Strategien eingehen und auch nicht auf die interessantere Frage, welche der *klassischen* Probleme durch den neuen Ansatz in welche neuen Fragestellungen transformiert werden. Ich möchte vielmehr das Augenmerk meiner Leser auf einen anderen Punkt lenken: Wenn wir mentale Repräsentate als von Systemen eingesetzte abstrakte Instrumente analysieren, dann werden unterschiedliche Typen von Instrumenten zur Instantiierung unterschiedlicher Typen von Intentionalität führen. Ich werde an dieser Stelle keine komplette Typologie mentaler Repräsentate anbieten können, sondern lediglich auf eine sehr einfache (und etwas idiosynkratische) Unterscheidung zwischen zwei Klassen von Repräsentaten hinweisen. Diese Unterscheidung scheint mir deshalb interessant, weil in ihr die Verknüpfung physischer und intentionaler Eigenschaften mentaler Repräsentate deutlich wird. Es ist die Unterscheidung zwischen „digitalen" und „analogen" Repräsentaten.

Propositionale Repräsentate sind digitale Repräsentate. Sie haben Wahrheitswerte und stehen untereinander in Folgerelationen. Der paradigmatische Fall solcher digitalen Codes sind natürliche Sprachen.[213] Digitale Codes zeichnen sich dadurch aus, daß die Relation zwischen ihren Elementen und den Repräsentanda willkürlich konstruiert ist. Die Konstruktion dieser Bedeutungsrelation ist unter anderem, wie wir von Wittgenstein gelernt haben, ein *soziales* Phänomen: Natürliche Sprachen entstehen in Gesellschaften. Die syntaktischen oder phonetischen Eigenschaften ihrer Elemente lassen keinerlei Rückschlüsse auf die Eigenschaften der jeweiligen Repräsentanda zu. Sätze *beschreiben* Sachverhalte, aber sie können sie nicht simulieren. Wahre Sätze besitzen eine Referenz, aber sie stehen in keiner wie auch immer gearteten Ähnlichkeitsrelation zu ihren Inhalten. Diese Eigenschaften machen sie ungeeignet für „mentale Testläufe", da solche Testläufe ja gerade ein komplexes Simulat erzeugen sollen, an dem

---

Diesem sind die einzelnen physischen Kausalketten und ihre Zeitabfolge intern nicht gegeben, sondern nur ihre Endprodukte. Nicht das Bewußtsein richtet sich auf die Welt, sondern komplexe physikalische Ereignisse determinieren mentale Modelle. Allerdings ist es richtig, daß die „Bewußtsein" erzeugende Metamodellierungsfunktion (vgl. Abschnitt 2.1.4) sich auf (naiv-realistisch „mißverstandene") mentale Modelle richtet.

[213] Obwohl es in natürlichen Sprachen meist auch Formen der Beschreibung durch Ähnlichkeit gibt, zum Beispiel durch onomatopoetische Weisen der Lautnachahmung.

die Eigenschaften eines Resultats möglicher Außenweltereignisse abgelesen werden können.[214]

Simulate dagegen haben die Funktion, die relationale Struktur ihrer Simulanda möglichst vollständig darzustellen (eben deshalb können sie nicht nur strukturelle Objekteigenschaften repräsentieren, sondern auch Prozeßabläufe). Der epistemische Prozeß also, der sich analoger Repräsentate bedient, hat als Ziel *Ähnlichkeit* und nicht Wahrheit. Man kann deshalb nicht oft genug betonen, daß „Wissen durch Sätze" und „Simulation durch interne Modelle" zwei grundverschiedene epistemische Phänomene sind, die zu anderen kognitionsphilosophischen Theorien führen. Welche internen Strukturen faktisch verantwortlich sind für menschliche Informationsverarbeitung hat also große Bedeutung für den Begriff der Intentionalität, der einer Theorie des Geistes zugrundegelegt werden muß. Propositionale und analoge Repräsentate als von informationsverarbeitenden Systemen eingesetzte Werkzeuge zur Verhaltenssteuerung erzeugen nämlich sehr unterschiedliche Formen der Bezugnahme auf die Realität. Colin McGinn betont ebenfalls diese Dichotomie, wenn er schreibt:

> . . . I have set the modelling theory in opposition to the sentential theory; neither type of theory is a special case of the other – though both may be species of a wider genus (internal structure theories). The reason for the opposition is that sentences do not simulate what they describe and models do not describe what they simulate. Simulating and describing are toto caelo distinct kinds of relations, calling for quite different features in the items they relate. Sentences are digital; models are analogue. Sentences have grammatical structure; models have the structure of the worldly things they simulate, which is not grammatical (unless these things happen to be themselves sentences). Sentences have semantic properties – truth conditions, reference, sense; models do not – any more than maps or tree rings do. Sentences and models reach out to reality using quite different kinds of limbs or hooks; in no way are the two theories notational variants of each other.[215]

Die Instantiierung von Wissen durch mentale Zustände wurde traditionell[216] als Relation zwischen einer Person und einer Proposition analysiert und unter der Überschrift „propositionale Einstellungen" von Philosophen des Geistes diskutiert. Eine der wichtigsten aus der Anwendung des Informationsverarbeitungsansatzes auf kognitive Leistungen resultierenden

---

[214] Johnson-Laird 1983 argumentiert für die These, daß sogar logische Operationen im Bereich des Mentalen auf mentaler Simulation beruhen: Konklusionen sind innerhalb eines durch die „Prämissen" gegebenen mentalen Modells der Welt erzeugte veränderte mentale Modelle. Vgl. hierzu auch Schnotz 1988: 312; Schnotz diskutiert empirische Befunde bezüglich der psychischen Realität von mit mentalen Modellen durchgeführten internen Simulationen. Er zieht Verbindungen zwischen einer prozeduralen Semantik und einer Konzeptualisierung von menschlichem Textverstehen als der Konstruktion von mentalen Teilmodellen. Vgl. Schnotz 1988: 323ff.

[215] Vgl. McGinn 1989: 181.

[216] Seit Russell 1940; vgl. auch Bieri 1987b: 25ff, Churchland, P.M.: 1981, 1986, Churchland, P.S. 1980, Dennett 1971, 1978, 1981, Stich 1983.

Einsichten lautet dagegen: Das Format der internen Repräsentate, zu deren Gehalt Personen in epistemischen Relationen stehen, spielt eine zentrale Rolle für die theoretische Analyse **a)** der epistemischen Relationen selbst, und **b)** der mit ihnen einhergehenden bzw. durch sie erzeugten psychologischen Eigenschaften. Eine Person, die Information in Form interner propositionaler Repräsentate besitzt, bezieht sich in einer völlig anderen Weise auf die Welt als eine Person, die Information in Form interner Simulate besitzt. Die erste Person mag sich in mentalen Zuständen des Typs „glauben, daß p" befinden und diese Zustände können als Einstellungen zu Propositionen mit Wahrheitswerten, Referenz usw. analysiert werden. Wie jedoch analysieren wir solche mentalen Zustände, die dadurch entstehen, daß Personen in Relationen zu komplexen Simulaten eintreten? Was unterscheidet analoges Wissen von digitalem Wissen, *Intentionalität durch Deskription* von *Intentionalität durch Simulation*?

> Der Unterschied zwischen einer propositionalen Repräsentation und einem mentalen Modell läßt sich seinerseits anhand einer Analogie verdeutlichen. Die Beziehung zwischen beiden entspricht der Beziehung zwischen einer Beschreibung und dem beschriebenen Sachverhalt bzw. in diesem Fall eines Analogmodells dieses Sachverhalts: Eine Beschreibung schränkt den Bereich der Möglichkeiten ein, ohne das Gemeinte vollständig zu determinieren. Sie ist eine partielle Spezifikation des Gemeinten (. . .). Ähnlich spezifiziert eine propositionale Repräsentation zwar das mentale Modell des im Text dargestellten Sachverhalts, ohne jedoch dieses bis ins einzelne festzulegen. Dieser Aspekt der Beziehung zwischen propositionalen Repräsentationen und mentalen Modellen ist auch vergleichbar mit der Beziehung zwischen Texten und Bildern. Bekanntlich kann ein Bild tausend Worte wert sein. Umgekehrt kann aber ein Text für eine unendliche Menge von Bildern stehen. Ähnlich kann der semantische Gehalt einer propositionalen Repräsentation durch eine Vielfalt von mentalen Modellen repräsentiert werden, die alle mit dieser propositionalen Repräsentation übereinstimmen.[217]

Auch Fred Dretske hat die epistemologischen Unterschiede zwischen analogen und digitalen Kodierungsmechanismen untersucht.[218] Er weist darauf hin, daß bei der Konvertierung von analogen Repräsentaten in digitale Repräsentate Information *verlorengeht*, während gleichzeitig eine Klassifikationsleistung (durch „Begriffsbildung") bezüglich der signifikanten Bereiche der Inputvariable erbracht wird.

> To describe a process in which a piece of information is converted from analog to digital form is to describe a process that necessarily involves the loss of information. Information is lost because we pass from a structure (. . .) of

[217] Vgl. Schnotz 1988: 311.

[218] Vgl. Dretske 1981, Kapitel 6. – David Lewis analysiert Analogrepräsentation von Zahlen als Repräsentation durch primitive oder quasi-primitive physikalische Größen, digitale Repräsentation von Zahlen dagegen als Repräsentation durch differenzierte „multidigitale" Größen, d. h. durch Größen, deren Werte arithmetisch von einer Reihe anderer „unidigitaler" Größen mit wenigen möglichen Werten abhängen. Vgl. Lewis 1971.

> greater informational content to one of lesser informational content. Digital conversion is a process in which irrelevant pieces of information are pruned away and discarded. Until information has been lost, or discarded, an information processing-system has failed to treat different things as essentially the same. It has failed to classify or categorize, failed to generalize, failed to „recognize" the input as being the instance (token) of a more general type.[219]

Konnektionistische Systeme sind in der Lage, auch durch die Erzeugung analoger Strukturen Generalisierungen und Begriffsbildungen vorzunehmen, indem sie Prototypen aktivieren.[220] Wenn das betreffende System ein menschliches Gehirn ist und der von ihm erzeugte Prototyp – zum Beispiel der Prototyp einer Katze – die Bedingungen erfüllt, die ihn zu einem potentiellen Bewußtseinsinhalt machen, dann kann man vom mentalen Modell des Begriffs der Katze sprechen. Das mentale Modell der Katze muß dabei keine kontextunabhängige, stabile Begriffsstruktur aufweisen, wie wir sie von natürlichen Sprachen und anderen externen Repräsentationssystemen kennen. Mentale Modelle sind Werkzeuge, die von Personen eingesetzt werden, um Wissen über die Welt zu erwerben und Handlungen zu steuern.[221] Welches innere Werkzeug eine Person in einer gegebenen Situation in sich erzeugt, hängt aber nicht nur von dem gegebenen Problem ab, sondern auch von ihrem internen Globalzustand: Wenn ein bestimmtes Modell sich problemlos in das aktuelle Modell der Welt einbetten läßt, ist dies für das System ein Indiz für hohen epistemischen Gehalt. Wenn die relationale Struktur eines mentalen Simulats sich ohne Schwierigkeiten in die relationale Gesamtstruktur des Weltmodells einbinden läßt, dann ist diese Entdeckung nämlich ein Anzeichen dafür, daß es sich um eine erfolgreiche, *funktionierende* interne Struktur handelt, die das Ziel der maximalen Ähnlichkeit oder Isomorphie mit ihrem Repräsentandum weitgehend erreicht. Was für satzartige Strukturen Wahrheit ist, heißt für von Biosystemen erzeugte Analogrepräsentate *funktionale Adäquatheit*.

Damit ist für menschliches, nicht-pathologisches Wachbewußtsein die erste Form von Wissen benannt: Mein Modell der Welt inkorporiert einen Teil meines Wissens über die Welt. Das bedeutet nicht, daß mir – als dem dieses Modell in sich erzeugenden physischen Gesamtsystem – dieses Wissen in Form aktuell *bewußter* mentaler Modelle gegeben sein muß. Vor allem bedeutet es nicht, daß mir dieses Wissen in Form einer mir selbst bewußten propositionalen Einstellung der Art „Ich weiß, daß p" mental gegeben sein muß. Wir erinnern uns: Mentale Modelle besitzen keine logische Form und keine Grammatik. Das bedeutet unter anderem, daß durch

---

[219] Vgl. Dretske 1981: 141.

[220] Vgl. zum Beispiel Churchland 1989, Kapitel 10.

[221] „Person" ist hier vorerst in einem Strawsonschen Sinne gemeint: Personen sind logisch primitive theoretische Entitäten, auf die sowohl psychologische als auch physikalische Prädikate angewendet werden können. Man darf sie nicht mit phänomenalen Personen verwechseln. „Phänomenale Personen" sind die mentalen Modelle, die Personen intern von sich selbst konstruieren. Vgl. Strawson 1972 und das nächste Kapitel.

mentale Modelle nicht-diskursives Wissen generiert wird. Dieses nicht-diskursive Wissen zeichnet sich dadurch aus, daß es durch ein erlebnismäßig nicht hintergehbares internes Modell der Welt dargestellt wird. Das Format der zur internen Darstellung verwendeten Repräsentate und Simulate ist ein analoges und das Kriterium, nach dem ihr epistemischer Gehalt vom System beurteilt wird, ist die Einbettbarkeit in einen repräsentationalen Gesamtzustand.

Propositionen befinden sich in logischen Räumen (und nicht in unseren Köpfen[222]) und die empirischen Belege für das Vorhandensein interner Satzanaloga in biologischen Gehirnen sind nicht besonders stark. Natürlich kann es mentale Modelle von Propositionen und Sätzen in öffentlichen Sprachen geben – in der Tat scheint das zentrale Spezifikum der typisch menschlichen Form von Kognition und Rationalität in der Interaktion von diskursivem Wissen durch externe, sozial generierte und digitale Repräsentationscodes und nicht-diskursivem, internem und durch Gehirnzustände erzeugtem, analogem Wissen zu sein. *Subjektives* Wissen jedoch ist immer Wissen außerhalb von Sätzen, es ist Wissen durch Analogizität und Ähnlichkeit und nicht durch Wahrheit und Falschheit.[223] Intersubjektives Wissen – etwa in Form wissenschaftlicher Theorien[224] – dagegen bedient sich externer Repräsentate in einem digitalen Format. Wären die zur Erzeugung subjektiven Wissens eingesetzten internen Strukturen auch Sätze, dann würde dies das Problem der Intentionalität bloß in Richtung auf das Projekt einer Semantik für die *Language of Thought* verschieben. An diesem Punkt wird die Überlegenheit der von mir in den vergangenen Abschnitten angedeuteten Theorie mentaler Repräsentation deutlich: Die zentralen Entitäten – mentale Modelle – sind bereits in einen kausal-teleologischen Kontext eingebettet, weil sie vom System benutzte Instrumente sind, die in sich eine Steuer-, eine Repräsentations- und eine Vorhersage- bzw. Simulationsfunktion vereinigen. Sie müssen nicht interpretiert werden, weil sie funktional aktive Strukturen[225] innerhalb eines Systems sind, welches bestimmte Probleme lösen und gewisse Ziele verfolgen muß. Ihr Gehalt ist ihr Gebrauch.

Eine sententialistische Theorie mentaler Repräsentation wie die Fodorsche hat auf der anderen Seite den Vorteil, mit wesentlich weniger struktu-

---

[222] Colin McGinn hat die Beziehung zwischen Propositionen (die Zustände der Welt beschreiben und durch sie individuiert werden) und mentalen Modellen (die Eigenschaften, Zustände und Prozesse der Welt simulieren sowie durch Gehirnprozesse aktiviert werden) als *Indikation* analysiert. Propositionen indizieren („aus dem logischen Raum heraus") diejenigen mentalen Modelle – *„states of the head"* – die propositionale mentale Zustände realisieren. Da Propositionen durch ihre Referenten individuiert werden, werden mentale Modelle *indirekt indiziert* durch Zustände der Welt. Vgl. hierzu McGinn 1989: 209ff.

[223] Subjektives propositionales Wissen könnte, wie bereits mehrfach angedeutet, in der mentalen Simulation von externen Symbolmanipulationsvorgängen bestehen.

[224] Auch Theorien kann man als interne Repräsentationen interpretieren: Das sie erzeugende System ist die Wissenschaftlergemeinschaft.

[225] Vgl. McGinn 1989, Kapitel 2.

rellen Annahmen (bezüglich Alphabeten, Transformationsregeln usw.) auszukommen. Dafür benötigt sie eine zusätzliche psychosemantische Theorie, um das Phänomen mentalen Gehalts *überhaupt* erklären zu können. Gegenüber dem philosophischen Problem der Intentionalität des Mentalen bleibt sie indifferent. Über diese Schwäche des linguistischen Ansatzes in der Erklärung des intentionalen Gehalts mentaler Zustände sagt McGinn:

> . . . it purchases the relative simplicity of its operating principles at the cost of poverty of explanatory power. By making those principles syntactic in nature it leaves all the semantic questions unanswered. The richness will then have to come from the semantic department. The bare idea of a sentential engine does not, as I have said, explain the nature of intentionality; it shunts that question into a mysterious shed labelled ‚semantics for the language of thought'. . . . sentence manipulation *by itself* does not do the necessary explanatory work, since sentences qua syntactic items do not have their semantics written on them. This makes for a stark contrast with the modelling theory, since models are *already* intrinsically semantically significant – they have the mechanism of intentionality engraved right on them. The sentential theory will have explained the distinctive features of content when it has given a semantic theory for the internal symbols it postulates; but the modelling theory explains these features just as it is, without supplemenation from a substantive semantics. . . . The modelling theory *requires* more of the brain because it *explains* more about the brain's capacities.[226]

Man kann also sagen: Funktional aktive Analogrepräsentate sind internen linguistischen Strukturen in der Erklärung von Wissenserwerb durch informationsverarbeitende Systeme und in Hinblick auf die Intentionalität mentaler Zustände an explanatorischer Potenz überlegen. Sie sind es deshalb, weil sie durch ihre relationale Struktur (und die Art ihrer physischen Realisierung) bereits in das kausal-funktionale Netzwerk eingebettet sind, in dem das System sich immer schon befindet.[227] Dieses Netzwerk bildet über den Selektionsdruck der biologischen Umwelt gleichzeitig seinen teleonomischen Hintergrund. Es ist letztlich dieser Hintergrund, der manche der internen Zustände biologischer Systeme zu solchen mit intentionalem Gehalt macht – zu einer Form *funktionalen Wissens.*

Ich hoffe, daß diese knappen Bemerkungen verdeutlichen konnten, daß das Format der von einem Repräsentationssystem intern eingesetzten

---

[226] Vgl. McGinn 1989: 202f.

[227] Dieses besondere „In-der-Welt-Sein" lebendiger Organismen, ihre Eingebettetheit in einen kausal-teleonomischen Kontext, markiert einen bedeutenden Unterschied zwischen natürlichen und künstlichen Systemen (so wie wir sie heute kennen). Künstliche Systeme haben keine *Interessen*, weil sie sich nicht in einem Überlebenskampf oder in einer Wettbewerbssituation (um Energiequellen, Fortpflanzungserfolg usw.) mit anderen Systemen befinden. Erst wenn wir künstliche Systeme auf geeignete Weise in einen kompetitiven *sozialen Kontext* einbetten, werden sie Verhaltensmuster entwickeln, die man extern als das Verfolgen von Absichten interpretieren kann. Wenn diese Systeme sich dann auch noch *intern* als Agenten mit spezifischen Interessen modellieren, besitzen sie das Potential, einmal zu phänomenalen intentionalen Systemen in einem genuinen Sinne zu werden.

Strukturen eine entscheidende Rolle spielt für die Art des von ihm hergestellten Weltbezugs und für den Charakter des so erzeugten Wissens über die Welt. Ich werde diesen Punkt und seine erkenntnistheoretischen Konsequenzen jedoch nicht weiterverfolgen, da ich in dieser Arbeit primär an dem von einem System hergestellten internen Selbstbezug und damit an einer durch mentale Modelle erzeugten psychologischen Eigenschaft interessiert bin: *Subjektivität*.

2.3 Gehirne als Generatoren von Modellen der Wirklichkeit

Für die Analyse phänomenalen Bewußtseins ist eine Eigenschaft der ihm zugrundeliegenden Strukturen von besonderer Bedeutung: Wechselseitige Einbettbarkeit durch rekursive Funktionen.[228] Mentale Modelle können ineinander eingebettet werden, sie können von Repräsentaten zu Repräsentanda werden und auf diese Weise als Bausteine höherstufiger psychischer Strukturen fungieren. Diese Tatsache verhilft uns vielleicht zu einer Erklärung der *prima facie* analyseresistenten Homogenität, Multimodalität und Bewußtheit phänomenaler Zustände. Zu diesen Punkten habe ich weiter oben bereits einige Bemerkungen gemacht. Die wechselseitige Einbettbarkeit mentaler Modelle besitzt jedoch außerdem Bedeutung für eine Theorie des Geistes, die uns erklären will, was es heißt, daß Gehirne *phänomenale Welten* erzeugen.

2.3.1 Repräsentationale Gesamtzustände

Endliche physikalisch oder biologisch zu beschreibende Automaten können zu einem gegebenen Zeitpunkt t nur eine endliche Anzahl interner Simulate konstruieren. Die Zahl der aktiven mentalen Modelle wird aufgrund der begrenzten Verarbeitungskapazität immer eine endliche sein. Diejenigen aktivierten Datenstrukturen, die untereinander verknüpft oder in eine höherstufige Struktur eingebunden sind, bilden zu einem gegebenen Zeitpunkt den jeweiligen *repräsentationalen Gesamtzustand* des Systems.[229] Der repräsentationale Gesamtzustand eines Systems besteht aus der Gesamtmenge der dem System aktuell *verfügbaren* Information über den Zustand der Welt. Darum kann man diesen repräsentationalen Gesamtzustand auch als das jeweilige „mentale Modell der Welt“ oder als das

[228] Vgl. Johnson-Laird 1983, Kap. 1.

[229] Daraus folgt, daß ein System mit ausreichender Verarbeitungskapazität theoretisch auch mehrere solcher Makro-Simulate parallel erzeugen kann. Diese Einsicht könnte eine Bedeutung für die Analyse und Interpretation von Schizophrenien und phänomenalen Spaltungszuständen im allgemeinen besitzen.

„mentale Modell der Wirklichkeit“ bezeichnen.[230] Für ein solches System sind die Grenzen seines internen Modells der Welt die Grenzen *seiner* Welt. In einem verschärften Sinn gilt dies noch einmal für Systeme mit phänomenalem Bewußtsein.

Das phänomenale Modell der Wirklichkeit besteht zu einem gegebenen Zeitpunkt immer aus all jenen Elementen des repräsentationalen Gesamtzustandes, die gerade Bewußtseinsinhalte darstellen. Wenn die in Abschnitt 2.1.4 skizzierte Hypothese bezüglich des „Bewußtheits-Aspekts“ von Subjektivität richtig ist, dann ist der bewußte Teil des Weltmodells immer genau derjenige, der durch eine spezifische, einheitliche Metarepräsentations-Funktion (die selbst nicht zum Repräsentandum werden kann) noch einmal mental modelliert wird. Wenn wir uns selbst als Erlebnissubjekte betrachten, so ist es genau diese – durch neuronale Informationsverarbeitung generierte – Sektion des mentalen Modells der Wirklichkeit, in der sich das psychische Drama unseres phänomenalen Lebens entfaltet. Auf die Frage, wer die Hauptrolle in diesem Drama spielt, werde ich im nächsten Kapitel eine neue Antwort zu geben versuchen.

Unser mentales Modell der Welt ist eine dynamische, sich permanent verändernde und multimodale Landkarte des uns zugänglichen Teilbereichs der Wirklichkeit. Je größer ihr Detailreichtum und ihre innere Konsistenz sind, desto *realer* erscheint sie dem Erlebnissubjekt phänomenaler Zustände.[231] Eine normale Landkarte, so wie sie im Klassenzimmer hängt, ist ein zweidimensionales, externes Analogrepräsentat. Sie verändert sich nicht mit der Landschaft, die sie repräsentiert, und sie benutzt nur eine einzige „Modalität“. Das von unserem Gehirn erzeugte phänomenale Modell der Welt auf der anderen Seite ist ein partiell räumliches Modell,

[230] Ich ignoriere hier systematisch eine Reihe möglicher theoretischer Komplikationen. Sie bestehen in der (logischen) Möglichkeit von a) „direktem“ Informationsgewinn über nichtphysikalische Trägermedien (außersinnliche Wahrnehmung, „klassische“ Intentionalität), b) Informationsgewinn durch noch unbekannte, naturwissenschaftlich zu beschreibende Medien, der nicht unter den Begriff der „Repräsentation“ fallen kann, und c) das Vorhandensein *multipler* Realitätsmodelle im System bzw. isolierter Modelle, die auf keine Weise in das aktuelle mentale Modell der Welt eingebunden und doch verhaltenswirksam sind.

[231] Diese These kann man durch Beobachtungen an Personen stützen, deren Welterfahrung zunehmend den Charakter der „Unwirklichkeit“ oder „Traumartigkeit“ trägt. (In schweren Fällen spricht die Psychiatrie von *Derealisation*). Wir alle kennen solche psychischen Zustände in schwächerer Ausprägung nach traumatischen Ereignissen verschiedenster Art (nach Unfällen, nach seelischen Schockerlebnissen): Der Detailreichtum unseres phänomenalen Weltmodells nimmt vorübergehend dramatisch ab und die Welt erhält eine „traumartige“ und „unwirkliche“ Qualität. Andererseits kann in solchen Zuständen, in denen die Verarbeitungskapazität und das allgemeine Erregungsniveau des Gehirns wahrscheinlich deutlich erhöht sind (zum Beispiel bei Manien, religiösen Erfahrungen, pharmakologisch induzierten Zuständen oder „Gipfelerlebnissen“) die phänomenale Wirklichkeit eine „überreale“ Qualität gewinnen. Diesen variablen phänomenalen Realitätsaspekt (sozusagen die subjektiv erlebte „Authentizität“ der Welt, die „Seinsgewißheit“) kann man hypothetisch mit der Informationsdichte des vom Gehirn erzeugten Weltmodells korrelieren.

welches intern die Replikation eines Teils der relationalen Struktur der Welt leistet.[232] Es verknüpft Information aus einer Vielzahl voneinander abgeschotteter Module und sein Gehalt ist einer permanenten Korrektur durch den jeweiligen Informationsfluß aus den Sinnesorganen und durch höhere kognitive Operationen unterworfen. Es besitzt außerdem eine Reihe „nicht-öffentlicher" und „qualitativer" Eigenschaften.

Man versteht besser, was es eigentlich bedeutet, daß Gehirne phänomenale Welten erzeugen, wenn man einen Blick auf pathologische oder *abweichende* phänomenale Realitätsmodelle wirft. Sicherlich haben einige der vorangegangenen Bemerkungen in den Ohren meiner Leser sehr abstrakt geklungen. Sie sollen aus diesem Grund durch einen Blick auf einige interessante faktische Bewußtseinszustände vor einer gewissen Blutleere bewahrt werden. Ein solcher Blick auf die Anomalien der psychischen Realität ist immer sinnvoll. Denn wenn man am besseren Verständnis eines komplexen Phänomenbereichs interessiert ist, hat eine Analyse von Grenzfällen und eingeschränkten Situationen sich häufig als heuristisch fruchtbar in bezug auf das Standardphänomen (in unserem Fall das normale Wachbewußtsein) erwiesen. Die Betrachtung von Grenzfällen komplexer Phänomene macht nämlich unbewußte Vorannahmen deutlich, klärt intuitive Fehlschlüsse auf und macht Defizite bestehender Theorien sichtbar. Mit diesen drei Zielen im Auge werde ich im folgenden Abschnitt einige solcher Fälle betrachten.

2.3.2 Deviante mentale Modelle der Welt: Agnosien, Neglekte, Blindsicht, Halluzinationen und Träume

Setzt man das menschliche Wachbewußtsein unter Standardbedingungen als Normfall, dann gibt es eine Reihe von Kriterien, nach denen man eine Systematisierung der fraglichen psychischen Phänomene erzielen kann. Man kann einen repräsentationalen Gesamtzustand nach seiner Funktionalität – *Schränkt er das innere und äußere Handlungssrepertoire einer Person ein? Erweitert er es?* – und nach seinem epistemischen Gehalt kategorisieren *(Ist seine Informationsdichte höher oder niedriger als die des Normalzustands? Beraubt er einen Menschen um spezifische Quellen inneren Wissens, erschließt er ihm neue?).* Man kann aber auch seinen phänomenalen Gehalt analysieren und dies ist in unserem Zusammenhang von Interesse. So könnte man fragen: *Welche phänomenalen Qualitäten gehen verloren, welche neuen treten hinzu? Wie ist die Qualität der Bewußtheit über das repräsentationale Spektrum verteilt – gibt es neue Bewußtseinsinhalte und/oder existieren Elemente des mentalen Modells der Welt, die von der metarepräsentierenden Funktion nicht mehr erfaßt werden können?*

[232] Die Arten von Relationen können dabei sehr unterschiedlicher Natur sein: räumlich, kausal, logisch, zeitlich usw.

Schließlich können über die Betrachtung struktureller Eigenschaften des phänomenalen Modells (interne Konsistenz, Speicherbarkeit, inhaltliche Verknüpfung mit früheren oder späteren Weltmodellen anderen Typs, Variabilität und Plastizität des Gehalts, Zentriertheit) Schlüsse auf die einer Instantiierung der interessierenden psychologischen Eigenschaften zugrundeliegenden Mechanismen der neuronalen Informationsverarbeitung (und damit auch die Formulierung neuer Hypothesen) vorbereitet werden.

Ich werde aber an dieser Stelle keines dieser Projekte verfolgen können, sondern mich mit einer kursorischen und unvollständigen Aufzählung einiger devianter mentaler Modelle der Welt begnügen müssen. Ich hoffe, daß auch durch diese knappen neurophänomenologischen Betrachtungen deutlich wird, was es heißt, daß mentale Modellierung unter anderem eine umfassende, phänomenale *Makro-Struktur* erzeugt – eine interne Struktur, die unsere subjektiv erlebte Welt ist. Die nun folgenden Beispiele haben alle eines miteinander gemein: Weder die vorwissenschaftlich-lebensweltliche Alltagspsychologie noch die klassischen philosophischen Theorien des Geistes können sie adäquat erklären.

(1) Agnosien

Können Sie sich vorstellen, wie es wäre, Ihr eigenes Gesicht nicht mehr im Spiegel wiedererkennen zu können? Ein Schlaganfall in einem bestimmten Bereich Ihres Gehirns könnte Sie dauerhaft in einen solchen *agnostischen* Zustand versetzen. Unter Agnosien leidende Patienten sind nicht mehr in der Lage, ihnen vormals bekannte Stimuli als solche wiederzuerkennen. Obwohl keinerlei Störungen der Wahrnehmungsfunktionen vorliegen und auch höhere kognitive Funktionen – einschließlich der Sprachverarbeitung[233] – nicht beeinträchtigt sind, kann der Patient die Bedeutung des jeweiligen Perzepts nicht mehr bewußt erfassen, weshalb diese Erkrankung früher auch als „Seelenblindheit“ bezeichnet wurde.[234] Die Ursache für solche Ausfälle sind häufig durch Schlaganfälle ausgelöste Schädigungen spezifischer Hirnbereiche. Im Falle visueller Agnosien sind dies zum Beispiel bilaterale Schädigungen derjenigen okzipito-temporalen Regionen, welche die unteren visuellen Assoziationsbereiche in Teilen der Brodmannschen Felder 18 und 19 betreffen. Eine phänomenologisch sehr eng umgrenzte und häufig zitierte visuelle Agnosie ist die eingangs erwähnte Prosopagnosie:[235] Sie besteht in der Unfähigkeit, Gesichter wiederzuerkennen – in manchen Fällen sogar einschließlich des *eigenen* Gesichts im Spiegel. Die visuelle Welt prosopagnostischer Patienten bleibt ansonsten uneingeschränkt – bis auf die Tatsache, daß sie einen ganz bestimmten Stimulus auf einer sehr spezifischen, kontextabhängigen Ebene nicht mehr mit der be-

[233] Es handelt sich hier also nicht bloß um Anomien, d. h. um die Unfähigkeit, Perzepte zu benennen.

[234] Vgl. Teuber 1965, Damasio 1987.

[235] Vgl. Damasio *et al.* 1982, Tranel/Damasio 1985. Vgl. auch Dennett 1991: Kapitel 11.

reits intern vorhandenen Wissensstruktur verknüpfen können. Patienten mit solchen Läsionen können ihre Freunde und Verwandten sofort und ohne Schwierigkeiten an der Stimme erkennen. Sie sind jedoch nicht mehr in der Lage, die Identität von Personen (einschließlich ihrer selbst) durch die innere Aktivierung eines bestimmten Elements in ihrem phänomenalen Modell der Welt festzustellen. Durch einen Defekt auf der Hardware-Ebene können also bestimmte *Integrationsleistungen* (nämlich die Einbettung eines bestimmten aktuellen Personenmodells in ein bereits bestehendes) nicht mehr erbracht werden. Das hat zur Folge, daß das interne Modell der Welt nicht nur funktional eingeschränkt, sondern auch phänomenal depriviert wird: Personale Identität *als visuell gegebene* verschwindet für immer aus der Erlebniswelt des Subjekts.

(2) Neglekte

Neglekte sind selektive Störungen der Aufmerksamkeit, d. h. *gerichteter Bewußtheit*.[236] Patienten mit den entsprechenden Syndromen können ihre Aufmerksamkeit nicht mehr auf bestimmte Sektionen des mentalen Modells der Wirklichkeit richten, was perzeptuelle, motorische und motivationale Folgen nach sich zieht.[237] Eine der bekanntesten Formen, der *Hemi-Neglekt*, tritt nach unilateralen Schädigungen oder Tumoren – überwiegend in der rechten Hemisphäre – auf. Manche solcher Patienten sind nicht mehr in der Lage, die linke Hälfte von Sätzen zu lesen oder Ereignisse in der linken Hälfte ihres Wahrnehmungsfeldes phänomenal zu repräsentieren. Viele von ihnen waschen oder kleiden ihre linke Körperhälfte nicht mehr, manche der männlichen Patienten hören auf, die linke Seite ihres Gesichts zu rasieren. Oliver Sacks beschreibt eine Patientin, die unter den Folgen einer Läsion in ihrer rechten Hemisphäre litt:

> Manchmal beschwert sie sich bei den Schwestern, sie hätten ihr keinen Kaffee oder Nachtisch auf ihr Tablett gestellt. Wenn sie dann antworten: „Aber Mrs. S. da steht es doch – links von Ihrem Teller", scheint sie nicht zu verstehen, was sie sagen, und sieht nicht nach links. Wenn man ihren Kopf sanft nach links dreht, so daß das Dessert in der intakten rechten Hälfte ihres Gesichtsfeldes erscheint, sagt sie: „Ach, da ist es ja – aber eben war es noch nicht da." Sie hat den Begriff „links", bezogen sowohl auf die Außenwelt als auch auf ihren eigenen Körper, vollständig verloren. Manchmal beklagt sie sich, ihre Portionen seien zu klein, aber das kommt daher, daß sie nur von der rechten Hälfte des Tellers ißt. Es kommt ihr nicht in den Sinn, daß er auch eine linke Hälfte

[236] Nicht-gerichtete Aufmerksamkeit oder *Vigilanz* scheint dagegen ein allgemeiner Hintergrundparameter unseres repräsentationalen Gesamtzustandes zu sein, der sich auf der phänomenalen Ebene als allgemeine „Bewußtseinshelligkeit" manifestiert. Diese psychische Qualität könnte Resultat einer generalisierten Form mentaler Präsentation (vgl. Abschnitt 2.1.3) sein, da sie einen über das gesamte phänomenale Modell der Welt verteilten Signalaspekt darstellt – der nicht-relational ist, aber in unterschiedlichen Intensitäten auftreten kann. Wir wissen, daß das wichtigste materielle Substrat für Vigilanz die Aktivität der *formatio reticularis* ist.

[237] Vgl. Mesulam 1987, Bisiach/Vallar 1988.

> hat. Manchmal trägt sie Lippenstift und Make-up auf – aber nur auf die rechte Seite ihres Gesichtes. Die linke läßt sie völlig unbeachtet. Eine Behandlung dieses Fehlverhaltens ist fast unmöglich, denn man kann ihre Aufmerksamkeit nicht darauf lenken (. . .) und sie ist sich keines Fehlers bewußt. Intellektuell weiß und versteht sie, daß sie etwas falsch macht, und lacht darüber, aber es ist ihr unmöglich, es direkt zu erfahren.[238]

Was die Patientin verloren hat, ist also gerade nicht der *Begriff* „links" oder die Fähigkeit zur Aktivierung gewisser propositionaler Repräsentate und linguistischer Funktionen. Was ihr fehlt, ist der Präsentationsaspekt, das mentale Präsentat ihres linksseitigen Wahrnehmungsfeldes. Unabhängig von allen kognitiven Operationen verschwindet damit auch der qualitative und der Bewußtheitsaspekt der linken Hälfte ihres mentalen Modells der Welt. Um diesen Verlust zu kompensieren, kann die Patientin jedoch andere Formen mentaler Repräsentation zu Hilfe nehmen:

> . . . Wenn ihr die Portionen zu klein vorkommen, dreht sie sich, die Augen nach rechts gerichtet, rechts herum, bis die „fehlende" Hälfte in ihr Blickfeld kommt; dann ißt sie diese, oder vielmehr die Hälfte davon, worauf sie sich weniger hungrig fühlt als zuvor. Aber wenn sie noch nicht satt ist oder über die Sache nachdenkt und den Eindruck gewinnt, sie könnte nur die Hälfte der fehlenden Hälfte gesehen haben, macht sie eine zweite Drehung, bis sie das verbleibende Viertel sieht, von dem sie wieder die Hälfte ißt. Das reicht gewöhnlich aus, um ihren Hunger zu stillen – immerhin hat sie jetzt sieben Achtel ihrer Mahlzeit verzehrt. Manchmal aber, wenn sie sehr hungrig oder besonders gründlich ist, führt sie noch eine dritte Drehung aus und ißt ein weiteres Sechzehntel ihrer Mahlzeit. (Dabei bleibt natürlich das linke Sechzehntel auf dem Teller unbemerkt.) „Es ist absurd", sagt sie. „Ich komme mir vor wie Zenons Pfeil: Ich erreiche nie das Ziel. Mag sein, daß es komisch aussieht, aber was bleibt mir unter diesen Umständen anderes übrig?"[239]

Daß die aus der Beobachterperspektive augenfällige *Diskontinuität* mentalen Gehalts subjektiv nicht erlebt wird, daß Abwesenheit von Information nicht dasselbe ist wie Information über eine Abwesenheit, wird von materialistischen Philosophen immer wieder betont: *„. . . . , the absence of representation is not the same as the representation of absence. And the representation of presence is not the same as the presence of representation."*[240] Auch Patricia Churchland hat darauf hingewiesen, daß wir alle naiverweise annehmen, visuelles Bewußtsein unterliege keinerlei räumlichen Begrenzungen[241] – wenn wir visuell bewußt sind, sind wir eben visuell bewußt. Wir erleben den nicht repräsentierten Teil der Welt hinter unserem Rücken nicht als ein phänomenales Loch in unserer subjektiven Wirklichkeit. Ähnlich wie vielen Neglekt-Patienten erscheint uns die Welt trotz dieses Defi-

[238] Vgl. Sacks 1987: 111.
[239] Vgl. Sacks 1987: 112.
[240] Vgl. Dennett 1991: 359.
[241] Vgl. Churchland 1988: 289.

zits als durchaus komplett. Philosophisch interessant ist deshalb an dem hier geschilderten pathologischen Fall, daß auch eine starke räumliche Einschränkung des phänomenalen Modells der Welt nicht mit einem phänomenalen Bewußtsein der Eingeschränktheit einhergehen muß.[242] Diese Tatsache zeigt einmal mehr, was *phänomenale Unhintergehbarkeit* heißt: Das Subjekt kann die phänomenale Welt nur propositional transzendieren – niemals phänomenal.

(3) Blindsicht

Von besonderem theoretischen Interesse sind Störungsbilder, die eine Dissoziation funktionaler und phänomenaler Zustände zur Folge haben. Damit sind all jene Situationen gemeint, in denen die funktionale Analyse eines Systems bzw. der interne Informationsfluß bis hin zum motorischen Output weitgehend unverändert bleibt, es auf der Ebene des phänomenalen Modells der Wirklichkeit jedoch zu Ausfällen oder Umstrukturierungen kommt. *Blindsicht* stellt ein solches Phänomen dar.

Bei Patienten mit einer Läsion innerhalb der geniculostriatalen Projektion des visuellen Cortex läßt sich ein Skotom nachweisen, ein „blinder Fleck" in dem korrespondierenden Bereich ihres Sehfeldes. In diesem Bereich des visuellen Modells befindet sich, so könnte man sagen, ein phänomenales Loch: Es gibt keine visuellen Bewußtseinsinhalte in bezug auf diesen Bereich[243] der Welt. Trotzdem können solche Patienten, wie sich experimentell und unter Einsatz nicht-verbaler Techniken zeigen läßt, komplexe visuelle Informationsverarbeitung durchführen. Sie können mit erstaunlichem Erfolg die Präsenz oder Abwesenheit von Zielobjekten „erraten" oder innerhalb des Skotoms präsentierte Farben und Muster unterscheiden. All diese Leistungen werden nach den Berichten der Patienten erbracht *ohne daß es zu einem subjektiven Seherlebnis kommt* – daher der Name dieser Störung. Manche der Versuchspersonen beschreiben die innere Erfahrung, die sie während erfolgreicher Versuche „nicht-erlebnisbegleiteten Sehens" durchleben, als Raten, andere dagegen protestieren sogar bei den Versuchsleitern gegen eine vermeintliche Aufforderung zur Lüge.[244]

Die philosophische Interpretation dieses empirischen Materials ist nicht einfach, zumal manche Patienten über „Ahnungen" oder diffuse emotionale Bewußtseinsinhalte vor einem erfolgreichen Akt des „Ratens" berichten. Sicher scheint, daß es in den fraglichen Bereichen zu weitreichender perzeptueller Verarbeitung bei gleichzeitiger Abwesenheit phänomenalen Bewußtseins kommt. Diese Möglichkeit einer Dissoziation von intentionalem und phänomenalem Gehalt des entsprechenden mentalen Modells

[242] Vgl. Bisiach 1988.

[243] Vgl. zum Beispiel Bodis-Wollner 1977, Cowey 1979, Cowey/Stoerig 1991b, Pöppel *et al.* 1973, Stoerig *et al.* 1985, Stoerig/Cowey 1990, 1991a, 1992, 1993 Weiskrantz *et al.* 1974, Weiskrantz 1986, Werth 1983, Zihl 1980.

[244] Vgl. Weiskrantz 1988: 188f.

ist ein klares Indiz für die Distinktheit der zugrundeliegenden Mechanismen. Solche empirischen Einsichten müssen ihren begrifflichen Niederschlag in unserer Theorie der mentalen Repräsentation finden.

Ist es nun die Abwesenheit des *Präsentationsaspekts*, der zur Erzeugung eines deprivierten mentalen Modells der Welt führt, oder löscht der physische Defekt eine Art Überwachungs- oder *Metarepräsentationsfunktion* im mentalen Leben der Patienten aus? Wenn es der Signalaspekt ist, der für den phänomenalen Aspekt der visuellen Leistung verantwortlich ist, dann könnte Blindsicht nur eine quantitativ abgeschwächte Form des normalen Sehens sein. In der Tat wird die Möglichkeit eines graduellen Übergangs von blindem Raten zu dem, was Weiskrantz als *non-veridical ‚seeing'*[245] bezeichnet, und vielleicht sogar bis hin zum „echten" Sehen von empirischen Forschern diskutiert. Die für eine philosophische Theorie des Geistes wichtige Frage ist in diesem Zusammenhang diejenige nach der faktischen Dissoziierbarkeit der Phänomenalität eines visuellen Modells und seiner repräsentationalen Struktur bzw. seines Gehalts. Ist die Bewußtheit unseres Sehens das Resultat einer gewissen minimalen Signalstärke der neuronalen Mechanismen, die das visuelle Modell aktivieren – oder gibt es eine diskrete Meta-Modellierungsfunktion, die bei Verletzungen der betreffenden Hirnregionen einfach ausfällt? Was genau fehlt einem Blindsicht-Patienten: Eine bestimmte Klasse mentaler Präsentate oder die diese Präsentate in bewußte Aspekte mentaler Zustände verwandelnde Meta-Repräsentationsfunktion? Die Tatsache, daß manche Blindsicht-Patienten – indem sie auch in ihrem „blinden Fleck" Farben unterscheiden können[246] – ein gutes Diskrimationsvermögen bezüglich qualitativen Gehalts demonstrieren, scheint eher auf die zweite Möglichkeit hinzudeuten.

Man sieht: Mehr als ein hypothetischer Mechanismus könnte für die Entstehung eines „phänomenalen Lochs" in einem mentalen Modell der Welt verantwortlich gemacht werden. Und die weitgehende Dissoziierbarkeit der funktionalen von den phänomenalen Eigenschaften des fraglichen mentalen Modells läßt vorsichtige Schlüsse auf die Distinktheit der zugrundeliegenden physischen Mechanismen zu. Dissoziative Störungen sind interessant für die Theoriebildung, weil sie uns die Möglichkeit von Bewußtseinskonfigurationen zeigen, die häufig unseren Intuitionen zuwiderlaufen. Unsere Intuitionen sind immer geprägt durch diejenigen mentalen Reali-

[245] Vgl. Weiskrantz 1988: 189.

[246] Vgl. Cowey/Stoerig 1991b, Pöppel 1987, Stoerig/Cowey 1990, 1991a, 1992, 1993; Dennett 1991: 326ff. Blindsicht als empirisch gut dokumentiertes Phänomen zeigt also, daß es etwas gibt, das viele Philosophen früher als *contradictio in adjecto* bezeichnet hätten: unbewußtes Farbensehen. Wir können uns das subjektive Erleben von Blindsicht nicht vorstellen, weil unsere Gehirne die entsprechenden mentalen *Simulate* nicht erzeugen können. *Logisch* möglich und empirisch plausibel sind nicht von einer Meta-Repräsentationsfunktion erfaßte mentale Präsentate auf jeden Fall. Eine ausgezeichnete Aufarbeitung der neuesten empirischen Erkenntnisse über die Bedingungen des Farbensehens, die speziell auf die philosophische Diskussion zugeschnitten ist, bietet Hardin 1988.

tätsmodelle, die wir im Laufe unseres Lebens kennengelernt haben. Man kann – in einem metaphorischen Sinn – mentale Realitätsmodelle als nicht-öffentliche Theorien über das Wesen der Wirklichkeit betrachten, die von Gehirnen intern mittels nicht-digitaler Repräsentationscodes erzeugt werden. Es ist deshalb wichtig, nicht stillschweigend Annahmen unserer Gehirne mit in unsere wissenschaftlichen Theorien über uns selbst und die Welt zu übernehmen. In der Philosophie des Geistes sind aus diesem Grund Gedankenexperimente besonders gefährlich und darum bedürfen ihre impliziten, häufig durch Introspektion gewonnenen Prämissen einer sehr kritischen Untersuchung.[247]

Das bizarre Spiegelbild von Blindsicht heißt *Antons Syndrom*.[248] Patienten, die durch eine Läsion des visuellen Cortex von plötzlicher Blindheit überfallen werden, bestehen in manchen Fällen hartnäckig darauf, *daß* sie noch sehen. Sie stoßen sich gleichzeitig an Möbelstücken und zeigen alle Anzeichen funktionalen Blindseins. Trotzdem verhalten sie sich so, als ob ihnen das subjektive Verschwinden der visuellen Welt nicht subjektiv *bewußt* ist; so produzieren sie zum Beispiel auf Fragen nach ihrer Umgebung falsche, aber konsistente Konfabulationen. Sie erzählen Geschichten über nicht-existente phänomenale Welten, die sie selber zu glauben scheinen, und streiten jeden funktionalen Defizit in bezug auf ihre Sehfähigkeit ab.

Ich möchte an dieser Stelle nicht die Frage der erkenntnistheoretischen Autorität des Subjekts über seine eigenen mentalen Zustände diskutieren[249], sondern die Aufmerksamkeit meiner Leser auf den interessanten Punkt lenken, daß wir es hier mit einer Art „rekursiven Neglekts“ oder einer „Meta-Agnosie“[250] zu tun haben: Das mentale Modell der Welt besitzt in diesen tragischen Fällen nämlich zwei „phänomenale Löcher“. Das **erste** erstreckt sich über die gesamte visuelle Modalität, das **zweite** ist ein phänomenales Loch *in bezug auf ein phänomenales Loch*. Es gibt also sogar phänomenale Defizite *zweiter Ordnung*, und die durch sie entstehenden Bewußtseinszustände falsifizieren endgültig und eindeutig die klassische cartesianische Annahme von der Selbsttransparenz des Bewußtseins. Unser Begriff phänomenalen Bewußtseins muß uns – um nicht mit empirischen Erkenntnissen zu kollidieren – eine Erklärung für Fälle geben, in denen phänomenale Ausfälle selbst nicht mehr phänomenal repräsentiert werden können. Meine eigene provisorische Hypothese für Fälle wie Antons Syndrom lautet: Der Verlust des visuellen Modells der *Welt* führt bei

---

[247] Vgl. Wilkes 1988, Kapitel 1.

[248] Vgl. Anton 1899, Benson/Greenberg 1969.

[249] Vgl. hierzu Rorty 1981a, 1981b; weitere Literaturangaben zur These der Inkorrigibilität mentalistischer Selbstzuschreibungen finden sich in Bieri 1981: 344f.

[250] Der korrekte Fachausdruck für Störungen, bei denen die bewußte Einsicht in eine Erkrankung fehlt, lautet „Anosognosie“. Sie entstehen durch nicht-dominante, parietale Läsionen. Damasio weist darauf hin, daß es sich bei ihnen nicht um Agnosien in einem engeren Sinne handelt. Vgl. Damasio 1987.

menschlichen Gehirnen nicht automatisch zu einer Änderung des *Selbstmodells*. Funktionale Blindheit kann gleichzeitig bestehen mit einem erlebnismäßig unhintergehbaren Modell des Selbst *als einer sehenden, visuell unversehrten Person* – möglicherweise ein Hinweis darauf, daß die zugrundeliegenden Mechanismen der Informationsverarbeitung weitgehend entkoppelt sind.

Bisher haben wir phänomenal eingeschränkte Realitätsmodelle betrachtet. Es gibt jedoch auch eine große Anzahl *hypertropher* Realitätsmodelle, d. h. repräsentationaler Gesamtzustände, die eine – meist unerwünschte und unkontrollierbare – Ausdehnung oder Erweiterung der phänomenalen Welt mit sich bringen.

(4) Halluzinationen

Für menschliche Systeme ist ein wichtiges Kriterium ihrer „psychischen Gesundheit" das Verhältnis von mentaler Repräsentation und mentaler Simulation. Eine Reihe devianter Bewußtseinszustände wird nun gerade dadurch ausgelöst, daß das System intern von einer großen Anzahl mentaler Simulate *überschwemmt* wird, was zu einer drastischen Verschiebung der Ratio von Repräsentation und Simulation in dem so enstehenden hypertrophen mentalen Modell der Welt führt. Überwiegend sind die in solchen Zuständen entstehenden mentalen Simulate reine Artefakte: Sie sind nicht Elemente eines kognitiven Prozesses im engeren Sinne, d. h. sie erfüllen keine Funktion *für* das System. Allerdings kann man andererseits gerade komplexe Halluzinationen und Wahngebilde als den verzweifelten Versuch eines mit unkontrollierbaren internen Signalquellen und repräsentationalen Artefakten konfrontierten Systems interpretieren, auch in einer solchen Situation noch ein möglichst konsistentes und gehaltvolles internes Modell der Welt zu konstruieren.

Eine wichtige Unterscheidung ist in diesem Zusammenhang diejenige zwischen *Halluzinationen* und *Pseudo-Halluzinationen*. Echte Halluzinationen sind unhintergehbar: Die meist akustischen Halluzinationen unter denen Schizophrene leiden, wenn sie Stimmen hören oder ihnen „fremde Gedanken gesendet" werden, sind möglicherweise vom auditorischen Cortex erzeugte Simulate, welche auf der phänomenalen Ebene *als Repräsentate* kodiert werden. Das System ist aufgrund eines pathologischen Gesamtzustands nicht mehr in der Lage, die Tatsache zu erkennen, daß die mentalen Modelle durch interne Ursachen aktiviert werden. Aus echten Halluzinationen kann sich das psychologische Subjekt unter anderem deswegen nicht mehr befreien, weil der Mechanismus, durch den die das System überflutenden Artefakte irrtümlicherweise als (durch die Standard-Kausalketten ausgelöste) Repräsentate interpretiert werden, seinerseits **erstens** nicht in Form eines bewußten mentalen Modells intern gegeben ist und **zweitens** durch höhere kognitive Operationen wie begriffliches Denken nicht penetriert werden kann. Das bösartige Flüstern höhnischer Geistwesen oder die Engelschöre aus höheren Ebenen der Wirklichkeit stellen für

den Schizophrenen eine aufgezwungene und unkontrollierbare Erweiterung seiner phänomenalen Welt dar, gegen die er sich *als phänomenales Subjekt* nicht wehren kann. Das ist deshalb so, weil das die phänomenale Welt erzeugende Gehirn partiell die Fähigkeit verloren hat, zwischen mentaler Simulation und ihrem eingeschränkten Sonderfall, mentaler Repräsentation, zu unterscheiden.

Auch wenn dem System diese Unterscheidung gelingt, kann es sich in einer Situation wiederfinden, in der Teile der für mentale Simulation verantwortlichen Hirnfunktionen außer Kontrolle geraten und temporär autonom geworden sind. In solchen Fällen kann das subjektiv erlebte Weltmodell durch eine Flut mehr oder weniger stabiler *phänomenaler Artefakte* angereichert werden. Sie können zwar vom System – auch auf der Ebene phänomenaler Modellierung – als afunktionale Simulate, man kann sagen: als überflüssige mentale Strukturen ohne funktionale Rolle in der psychischen Ökonomie des Systems, erkannt werden. Trotzdem sind auch Pseudo-Halluzinationen (die vom psychologischen Subjekt als solche erkannt werden) in ihrem phänomenalen Gehalt unhintergehbar.

Werden Halluzinationen durch pharmakologische Stimuli ausgelöst (etwa durch die Einnahme von LSD, Meskalin oder DMT), dann kommt es in unterschiedlichen Hirnbereichen zu einer weitgehend unspezifischen Enthemmung der neuronalen Aktivität.[251] Dies führt unter anderem dazu, daß der pure Signalaspekt der Verarbeitung unter Umständen extrem verstärkt wird – was auf der phänomenalen Ebene zur Intensivierung vieler Formen von qualitativem Gehalt führt. Gleichzeitig erhöht sich die Geschwindigkeit höherer mentaler Operationen bis hin zu Gedankenflucht und Desorientiertheit. Nimmt die Zahl und die Aktivierungsgeschwindigkeit der das System überschwemmenden mentalen Simulate so stark zu, daß es sie nicht mehr zu einem einzigen Modell der Welt und des Selbst *integrieren* kann (indem es die verschiedenen Halluzinationen zu einer „Geschichte" zusammenfaßt), dann kann es zu phänomenalen Spaltungszuständen kommen, welche wir üblicherweise als „psychotisch" zu bezeichnen pflegen. In solchen Fällen könnte man sagen, daß das mentale Realitätsmodell mit soviel phänomenalem Gehalt überfrachtet worden ist, daß es sich aufzulösen oder zu spalten beginnt.

Ein wichtiger Punkt scheint in diesem Zusammenhang zu sein, daß phänomenale Hypertrophie nicht gleichbedeutend ist mit epistemischer Expansion. Wenn die in ein Realitätsmodell eindringenden zusätzlichen mentalen Simulate, die wir „Halluzinationen" nennen, tatsächlich afunktionale Artefakte sind[252], dann wird das System als Ganzes zwar um phänomenalen

[251] Vgl. Aghajanian *et al.* 1968, 1970, 1975; Jacobs/Trulson 1979; Siegel 1975.

[252] Diese Frage muß von Einzelfall zu Einzelfall sorgfältig und differenziert entschieden werden. Die moderne halluzinogen-unterstützte Psychotherapie zum Beispiel hat gute katamnestische Resultate vorzuweisen, was als ein Indiz dafür gelten kann, daß es sich hier nicht *nur* um die Auslösung epistemisch leerer Prozesse handelt. Vgl. Leuner 1981.

Gehalt bereichert – aber nicht um *Wissen über die Welt*. Gibt es Zustände, die epistemisch *völlig* leer sind, aber einen reichen subjektiven Gehalt besitzen?

(5) Träume

Träume sind nicht-pathologische, episodisch wiederkehrende repräsentationale Gesamtzustände von Menschen, die sich durch eine ganze Reihe von Kriterien vom üblichen Wachbewußtsein unterscheiden. Die phänomenale Welt des Traums ist wesentlich instabiler als die des Wachzustandes. Die mentalen Modelle, aus denen sie sich zusammensetzt, scheinen tatsächlich anomische Entitäten zu sein, was ihre Beziehung zu anderen Elementen der phänomenalen Ebene angeht.[253] Die Inhalte unseres Traumbewußtseins ändern sich ständig, unvorhersehbar und nicht selten auf bizarre Weise. Emotionen können stark intensiviert sein, während auf der anderen Seite Schmerzwahrnehmungen sowie Geruchs- und Geschmackserlebnisse seltener auftreten als im Wachzustand.[254] Der Traum ist außerdem ein Beispiel für Bewußtsein ohne Erinnerung: Sogar beim Aufwachen gelingt es uns nur teilweise, phänomenale Fragmente aus der Traumwelt im Kurzzeitgedächtnis lange genug aufrechtzuerhalten um sie – via innerer oder äußerer Berichterstattung – festzuhalten und in re-präsentierbare Episoden unserer psychischen Biographie zu verwandeln. Intern dagegen können Träume sogar *hypermnestische* Zustände sein. Denn häufig sind Erinnerungen in unser Traumerleben eingewoben, die uns im Wachzustand nur sehr schwer zugänglich gewesen wären. Eigenschaften wie deutliche kognitive Fehlleistungen, komplexe Halluzinationen, Amnesien und Überemotionalisierungen machen den Traum zu einem Zustand, der sich auch als neurobiologisches Modell für andere veränderte Bewußtseinzustände (wie etwa die Schizophrenie) eignet.

Für eine philosophische Oneirologie sind zwei Aspekte des Traumzustandes von besonderem Interesse. **Erstens** zeichnen sich Träume durch ein ganz bestimmtes *metakognitives Defizit* aus: In normalen Träumen fehlt dem psychologischen Subjekt jede Einsicht über das Wesen des Zustandes. Das bedeutet, daß der repräsentationale Gesamtzustand *selbst* nicht noch einmal phänomenal als zu einer bestimmten Klasse gehörig repräsentiert wird. (Interessanterweise gibt es auch solche Träume, in denen dieses metakognitive Defizit aufgehoben ist. In solchen Träumen *weiß* das Subjekt, in welchem Bewußtseinszustand es sich befindet. Da metakognitive Übergänge dieser Art von zentralem Interesse für eine Theorie des Subjektivität sind, werde ich im nächsten Kapitel auf sie zurückzukommen haben.) Die Selbsttransparenz unseres Bewußtseins ist im Traumzustand gering: Wir gleichen Betrunkenen, die sich in eine Geisterbahn verirrt haben. Stärkste

[253] Bezüglich der Anomalie des Mentalen in Hinsicht auf die Formulierung strikter extramentaler, also psychophysischer Gesetze vgl. Davidson 1970, 1981.
[254] Vgl. Hobson 1987.

Inkongruenzen und Diskontinuitäten im Lauf der Dinge lösen keinerlei Zweifel in uns aus, wir geraten niemals in ein produktives, kritisches Verhältnis zur Realität und uns selbst – ein Zustand, der zumindest allen Philosophen verabscheuungswürdig erscheinen sollte. Unser Defizit ist jedoch **zweitens** nicht nur eines von „Meta-Bewußtheit" oder Zustandsklarheit, sondern auch ein epistemisches. Es gibt nämlich gute Gründe zu der Annahme, daß die mentalen Realitätsmodelle, die wir aus der Perspektive des Wachzustandes als unsere Träume zu bezeichnen pflegen, epistemisch weitgehend leer[255] sind. Wenn sich das aber als wahr herausstellen sollte, könnte man sagen: Träume basieren vollständig auf Simulation, es handelt sich bei ihnen aufgrund der Abwesenheit mentaler Repräsentate nicht um repräsentationale, sondern um *simulationale* Gesamtzustände. Ein kurzer Blick auf die funktionale Analyse und die neurobiologischen Grundlagen des Traumzustandes wird diese Vermutung jedoch nur teilweise bestätigen.

Die wichtigsten funktionalen Merkmale des Realitätsmodells „Traum" sind die folgenden:

(1) Output-Blockade: Träumer sind keine Handelnden. Das menschliche Gehirn ist nämlich, wenn es sich in dem für die Erzeugung des Realitätsmodells „Traum" notwendigen Zustand befindet, nicht in der Lage, motorischen Output zu erzeugen. Träumende Systeme sind aufgrund physischer Rahmenbedingungen nicht fähig, komplexe Verhaltensweisen oder Handlungen in der physischen Welt zu initiieren. Allerdings gibt es spezifische Formen von Mikro-Verhalten, wie zum Beispiel die schnellen Augenbewegungen *(rapid eye movements),* welche typischerweise die Traumphasen begleiten und nach denen diese Phasen auch als REM-Schlaf benannt wurden.[256] Träume sind somit Modelle der Welt, die in bezug auf die Verhaltenssteuerung funktionslos sind. Diese behaviorale Afunktionalität könnte jedoch einen tieferen Sinn haben: Sie verhindert, daß das System sich durch reafferente Signale selbst *weckt* – und so das gerade aktive mentale Modell der Wirklichkeit zum Zusammenbruch bringt. Der Schlafzustand selbst benötigt bestimmte physische Randbedingungen (zum Beispiel eine weitgehende Input-Blockade) und das System darf diese Randbedingungen nicht durch die „unabsichtliche" Ausführung von Motorbefehlen zerstö-

[255] Die epistemische Leerheit von Träumen ließe sich, wenn sie denn ein etabliertes Faktum sein sollte, trotzdem mit einer sehr hohen funktionalen Adäquatheit der betreffenden inneren Vorgänge für das System vereinbaren. Nach der Hypothese von Crick und Mitchison könnte der REM-Schlaf dem aktiven Verlernen „parasitischer Repräsentate" dienen. Auch Hobson und McCarley haben ihre ursprüngliche Hypothese vom „Zufallsgenerator" auf der Ebene des Hirnstamms mittlerweile zurückgenommen bzw. modifiziert. Eine weitere neue Hypothese zur entwicklungsgeschichtlichen Genese des REM-Schlafes und seiner Rolle für Gedächtnis und Reifung des Gehirns schlägt Winson 1991 vor.

[256] In Grenzfällen einer unvollständigen motorischen Blockade kann es auch – etwa beim „Sprechen im Schlaf" – zu nicht-intendierten komplexen Verhaltensformen kommen. Hier handelt es sich nicht um Handlungen, sondern um (wahrscheinlich afunktionale, d. h. keine Funktion *für* das System erfüllende) „behaviorale Artefakte".

ren. Die neurobiologischen Grundlagen dieser funktionalen Situation sind klar nachweisbar, sie bestehen in einer postsynaptischen Hemmung im letzten Bereich des gemeinsamen Pfads der Motorneuronen im Hirnstamm und im Rückenmark.[257]

(2) Input-Blockade: Träume sind Zustände, in denen periphere Signale nur selten zu den zentralen Mechanismen der Informationsverarbeitung durchdringen können. Träumen ist also kein Vorgang der normalen Makro-Repräsentation, weil seine Repräsentanda nicht aus überwiegend externen Ereignissen oder Situationen bestehen, die mentale Repräsentate auf dem Weg über die Standard-Kausalketten aktivieren. Deswegen sind Träume epistemisch leer zumindest in bezug auf den jeweils aktuellen Zustand der Systemumwelt – der ihnen zugrundeliegende Informationsfluß ist ein ausschließlich interner. Es gibt zwei gute Hypothesen bezüglich der neurobiologischen Realisierung dieser funktionalen Eigenschaft im menschlichen Gehirn, die von einer präsynaptischen Inhibition bestimmter afferenter Nerventerminals sowie gewisser Nuclei im Hirnstamm bzw. im Thalamus auf der einen Seite und von einer „Verstopfung" bzw. „Überschwemmung" höherer sensorischer Schaltkreise durch intern erzeugte mentale Simulate auf der anderen Seite ausgehen.

3) Interne Signalerzeugung: Das träumende Gehirn verarbeitet selbsterzeugte Eigenreize wie externen Input. Die philosophisch interessante Frage ist hier, ob diese Eigenreizung eine *Informationsquelle* im engeren Sinne darstellt – oder nur eine Sequenz von Zufallsereignissen auf der physischen Ebene. Ein guter vorläufiger Kandidat für die dem Gehalt von Traumerlebnissen zugrundeliegende Signalquelle sind vielleicht die sogenannten „PGO-Wellen". Im fraglichen Bereich des Hirnstamms gibt es nämlich eine gegenseitige Wechselwirkung zwischen aminergen und cholinergen Neuronen. Der Beginn von Traumphasen wird durch ein periodisches Zurruhekommen der aminergen Systeme eingeleitet, was wiederum zu einer Enthemmung funktional mit ihnen verknüpfter Einheiten und dann zum Entstehen der PGO-Wellen in der pontinen *Formatio reticularis* führt. Diese sich in den Thalamus und von dort in die visuellen und Assoziationscortices fortpflanzenden starken Impulse führen unter anderem zu nachweisbaren, geordneten und deutlich koordinierten Erregungsmustern in den okulomotorischen, vestibulären und visuell-sensorischen Zentren.

Interessant ist dabei, daß die intern erzeugte Information zumindest eine starke Raumspezifizität besitzt: Die den PGO-Wellen zugrundeliegende Zellaktivität bildet die gerichtete Aktivität der Augenbewegungen auf der Ebene des Hirnstamms noch einmal ab. Träume sind also nicht völlig sinnlos, ihre Repräsentanda sind geordnete interne Systemprozesse. Auch wenn diese Prozesse nicht als solche erkannt werden und auf der phänome-

[257] Hier und im folgenden stütze ich mich, was die empirischen Daten angeht, hauptsächlich auf Hobson 1987 und Hobson 1977.

nalen Ebene bloß eine bizarre Verkettung mentaler Simulate entsteht, so ist es wohl doch nicht die pure Aktivität eines „internen Zufallsgenerators“[258], die durch den Prozeß des Träumens mental modelliert wird. Und auch die *Interpretationsmechanismen*, die unser Gehirn einsetzt, um in der Konfrontation mit einer internen Signalquelle noch ein möglichst konsistentes Weltmodell zu erzeugen, tragen Information – zum Beispiel über die Persönlichkeitsstruktur des Träumers. In diesem Sinne könnte man sagen: Träume sind multidimensionale Rorschach-Tests, in deren Verlauf das Gehirn des Träumenden selbsterzeugte Zufallsfiguren zu einer komplexen Erzählung zusammensetzt, zu einem „inneren Märchen“. In Form einer Kette von konkreten subjektiven Erlebnissen manifestiert sich in diesem Märchen die Geschichte und die aktuelle Konfiguration des Systems, die Art und Weise in der es die Welt deutet. Träume sind also nicht epistemisch vollkommen leere Artefakte ohne biologische Funktion, sondern ein ganz bestimmter Typ von exklusiv *internem* Realitätsmodell, das nicht *als solches* erkannt wird. Sie werden durch ein episodisch wiederkehrendes physisches Ereignis (die Enthemmung sensomotorischer Schaltkreise als Folge einer Deaktivierung inhibitorischer, aminerger Neuronen) ausgelöst und zeichnen sich gegenüber dem Wachzustand durch eine Reihe kognitiver Defizite aus. Das theoretisch interessanteste dieser Defizite ist ein metakognitives: Das entstehende Realitätsmodell wird vom System selbst irrtümlicherweise in die Standard-Kategorie „Wachzustand“ eingeordnet. Das führt zum Ausfall einer wichtigen psychologischen Eigenschaft, nämlich der *Zustandsklarheit*: Träumer *wissen* nicht, daß sie Träumer sind.

Beenden wir hier unseren kurzen Blick auf abweichende Formen der mentalen Weltmodellierung. Ich hoffe, daß nun etwas klarer geworden ist, was ein mentales Modell der Wirklichkeit ist. Wir sollten jetzt zu unserem zentralen Thema zurückkehren indem wir uns der Frage zuwenden, welche Klasse von Realitätsmodellen es ist, die ein sie erzeugendes informationsverarbeitendes System zu einem System mit subjektiven Zuständen machen.

2.3.3 Monozentrierte Weltmodelle

Mentale Modelle der Welt sind die größten und umfassendsten internen Datenstrukturen, die wir kennen. Ich habe bereits darauf hingewiesen, daß eines der wichtigsten Desiderate für jede auf dem Begriff der mentalen Repräsentation aufbauende philosophische Theorie des Geistes eine Typologie mentaler Modelle sein wird. Eine solche Typologie wird mentale Universalien begrifflich als *Klassen von mentalen Modellen* analysieren und uns Auskunft darüber geben müssen, ob es spezies- und realisierungsunabhängige Arten von mentalem Gehalt gibt, auf deren Grundlage wir

[258] Vgl. Crick/Mitchison 1983, aber auch neuere Erkenntnisse über die funktionale Rolle des REM-Schlafes in Winson 1991.

eine universelle Psychologie entwickeln könnten. Auf der höchsten Stufe einer repräsentationalen Theorie des Geistes kann man in diesem Zusammenhang das folgende Projekt formulieren: Wir benötigen eine phänomenologisch, empirisch und begrifflich plausible Typologie mentaler Modelle der Wirklichkeit. Denn der Reichtum der dem menschlichen Gehirn zugänglichen internen Weltkonstruktionen ist kaum zu überschätzen – funktionale Komplexität erzeugt phänomenale Komplexität.[259] Wenn die Übergänge zwischen den vielen möglichen mentalen Wirklichkeitsmodellen nicht fließend sind und es sich herausstellen sollte, daß es *diskrete Typen* von repräsentationalen Gesamtzuständen gibt, dann könnte dieses Faktum das Fundament für eine naturalistische Theorie über die einem System möglichen Bewußtseinszustände bilden. Diskrete mentale Weltmodelle sind diskrete *Bewußtseinszustände.*

Eines der lohnendsten theoretischen Unternehmen in diesem Zusammenhang wird es sein, die strukturellen Kriterien zu analysieren, anhand derer wir verschiedene Klassen von repräsentationalen Gesamtzuständen voneinander unterscheiden können. Eine solche Analyse würde uns für ein gegebenes System die Palette und die Varianz der ihm zugänglichen internen Realitätsmodelle liefern – sie würde uns eine naturalistische Theorie seines phänomenalen Spektrums anbieten. Ich werde dieses Projekt hier nicht verfolgen, weil ich nur an einer ganz bestimmten Teilmenge mentaler Weltmodelle interessiert bin. Es ist jene Klasse von internen Modellen der Wirklichkeit, die mit Subjektivität eine theoretisch schwer zu fassende psychologische Eigenschaft instantiieren.

Wir erinnern uns: Die drei phänomenologischen Säulen des Subjektivitätsproblems sind der qualitative Gehalt und die Bewußtheit phänomenaler Zustände, sowie die Tatsache, daß sie an ein Selbst gebunden zu sein scheinen. Aus dem bisher Gesagten mag vielleicht hervorgehen, wie man das Problem des qualitativen Gehalts über den spezifischen Signalaspekt unterschiedlicher mentaler Modelle (d. h. durch ihre Unterlegung mit Präsentaten) lösen könnte und die Frage nach der Bewußtheit durch eine Analyse metarepräsentationaler Prozesse. Der Kern des philosophischen Problems besteht jedoch in der phänomenalen „Meinigkeit" – einer philosophisch sehr interessanten psychologischen Eigenschaft, die nur durch eine ganz bestimmte Klasse mentaler Weltmodelle instantiiert wird. Diese Klasse ist die Klasse der *zentrierten Weltmodelle.* Sie weisen ein strukturelles Merkmal auf, das die Charakteristik unserer inneren Erfahrung entscheidend prägt. Was das bedeuten soll, werde ich in dem nun folgenden Kapitel zu erläutern versuchen.

[259] Churchland (1989: 131) macht eine amüsante Milchmädchenrechnung bezüglich der Anzahl der dem menschlichen Gehirn möglichen internen Konfigurationen auf (nämlich $10^{100\,000\,000\,000\,000}$; im Vergleich dazu: die Zahl der Elementarteilchen im Universum wird auf 10^{87} geschätzt).

3. Kapitel

Die Selbstmodell-Theorie der Subjektivität

In den beiden vorangegangenen Kapiteln habe ich versucht, die Geographie der mit der Frage nach der Subjektivität mentaler Zustände verknüpften Problemlandschaft zu erkunden und den Hintergrund zu beleuchten, vor dem eine naturalistische Theorie des Geistes sich dieser Frage unter Einsatz des Begriffs der mentalen Repräsentation zuwenden kann. In den folgenden Kapiteln werde ich nun die bisher gewonnenen Einsichten auf die Frage mentaler *Selbstrepräsentation* anwenden und dann den Gang der Überlegungen um einige zentrale Elemente der philosophischen Gegenwartsdiskussion erweitern. Abschliessend werde ich versuchen, Antworten auf die zu Beginn aufgeworfenen Fragen zu geben. Wenden wir uns aber zunächst der Frage zu, was geschieht, wenn ein informationsverarbeitendes System sich selbst intern darzustellen beginnt.

3.1.1 Mentale Selbstrepräsentation: Mereologische Intentionalität

Mentale Selbstrepräsentation ist ein interessanter Sonderfall mentaler Repräsentation: Es ist genau der Fall, bei dem das den Prozeß der internen Repräsentation durchführende System sich *selbst* abbildet. Es erzeugt einen inneren Zustand – ein *Selbstrepräsentat* – mit dessen Hilfe es sich selbst auf nicht-sprachliche Weise beschreibt, mit dem es sein Verhalten steuern und auf sich selbst *als Ganzes* Bezug nehmen kann. Das Repräsentandum ist in diesem Fall das das mentale Selbstrepräsentat erzeugende System selbst. Die Realisierung dieser speziellen Variante der Repräsentationsbeziehung bringt einige neue logische und psychologische Eigenschaften mit sich. Um diese Eigenschaften besser zu verstehen, ist es gut, sich noch einmal das Grundschema der teleofunktionalistischen Abbildungsrelation zu vergegenwärtigen.

Mentale Repräsentation: $\mathbf{M}_{\mathbf{Rep}}$ (S,X,Y)
- X repräsentiert Y *für* S.
- X ist ein interner Systemzustand.
- X ist potentiell introspizierbar; d. h. kann seinerseits zum Repräsentandum von Repräsentationsprozessen höherer Ordnung werden.
- $\mathbf{M}_{\mathbf{Rep}}$ ist antisymmetrisch.

Betrachten wir zum Vergleich die logische Struktur des eben angesprochenen Sonderfalls:

Mentale Selbstrepräsentation: $\mathbf{M}_{\mathbf{S\text{-}Rep}}$ (S,X,S)
- X repräsentiert S *für* S.
- X ist ein interner Systemzustand.
- X ist potentiell introspizierbar; d. h. kann seinerseits zum Repräsentandum höherstufiger Repräsentationsprozesse werden.
- $\mathbf{M}_{\mathbf{S\text{-}Rep}}$ ist antisymmetrisch.

Mental sind diejenigen Bereiche eines internen Selbstrepräsentats, die potentielle Inhalte von phänomenalem Bewußtsein darstellen. Mentale Selbstrepräsentate sind komplexe interne Datenstrukturen in informationsverarbeitenden Systemen einer bestimmten Klasse. Diejenigen Teile solcher Datenstrukturen, die aktuelle Repräsentanda von höherstufigen mentalen Repräsentationsvorgängen sind, bilden die Grundlage des *bewußten* Modells, das ein System von sich selbst besitzt.[1]

Hier gilt vieles, was bereits in Abschnitt 2.1. über die Relation $\mathbf{M}_{\mathbf{Rep}}$ im allgemeinen gesagt worden ist. Die Relation $\mathbf{M}_{\mathbf{S\text{-}Rep}}$ hat ebenfalls eine Reihe von interessanten Eigenschaften. Das teleologische Zusatzkriterium wird (in unserem eigenen Fall) durch ein System mit biologischen Interessen näher spezifiziert: Mentale Modelle sind Instrumente *für* ein System. Zuallererst muß man also verstehen, daß – aus der Perspektive des Naturalismus – dieses System in einen Kontext eingebunden ist, der durch den erbarmungslosen Selektionsdruck einer biologischen Umwelt gebildet wird. Unter anderem durch diesen Kontext erhalten interne Simulate ihre Bedeutung, weil sie vom System innerhalb dieses Kontextes *benutzt* werden und als abstrakte Organe für das System fungieren. Auch Selbstrepräsentate sind solche Organe. Auf ihre Funktion werde ich später noch zu sprechen kommen.

Das System mit all seinen Interessen wird also gleichzeitig zum Objekt einer internen Abbildung, die ihrerseits diesen Interessen dient. Wenn diese Abbildung gut ist, ermöglicht sie dem System eine neue Klasse von inneren und äußeren Operationen: Das bewußte Verfolgen von Absichten.[2]

[1] Bei der Analyse von Selbstrepräsentaten als potentiellen Repräsentanda höherstufiger Darstellungsvorgänge ist zu beachten, daß Selbstrepräsentate in solchen Systemen, die die Fähigkeit zur inneren Selbstabbildung besitzen, nun natürlich auch auf mindestens zwei grundverschiedene Arten in mentale Meta-Modellierungen eingehen können: Sie können nämlich als Repräsentanda an der physischen Realisierung der Relationen $\mathbf{M}_{\mathbf{Rep}}$ oder $\mathbf{M}_{\mathbf{S\text{-}Rep}}$ beteiligt sein („Weltbewußtsein" oder „Selbstbewußtsein"). Repräsentate, deren Gehalt durch Eigenschaften des sie konstruierenden Systems gebildet wird, können „introspektiv" in Selbstrepräsentate eingebunden werden oder auch durch „gewöhnliche" Meta-Repräsentation zu Teilen eines Weltmodells werden („Projektion"). Die Tatsache, daß sowohl im Sinne der Relationen $\mathbf{M}_{\mathbf{Rep}}$ und $\mathbf{M}_{\mathbf{Sim}}$, als auch $\mathbf{M}_{\mathbf{S\text{-}Rep}}$ und $\mathbf{M}_{\mathbf{S\text{-}Sim}}$ höherstufige Formen mentalen Gehalts erzeugt werden können, schafft logische Komplikationen, neue psychologische Eigenschaften und ist unerläßlich für ein Verständnis von Nagels *objective self*.

[2] In einer naturalistischen Theorie des Geistes läßt sich die Verknüpfung der zwei semantischen Hauptelemente des Begriffs „Intentionalität" (die aus seiner lateinisch-arabischen Geschichte herrühren) als theoretisches und praktisches Intendieren über zwei Formen des Han-

durch diesen Zusammenhang wird ein System zu einem *Agenten*, es wird zum potentiellen Objekt ethischer Theorien, und zwar als mögliches moralisches Subjekt. Wenn es in der Lage ist, auch andere Systeme als Agenten mit Interessen und Absichten mental zu modellieren, dann tritt es nicht nur in einen sozialen Zusammenhang ein – es wird nun auch auf der psychologischen Ebene zum gesellschaftlichen Subjekt.

Ein **zweites** Merkmal des natürlichen Phänomens innerer Selbstbeobachtung, das ich durch die Relation $\mathbf{M}_{\mathbf{S\text{-}Rep}}$ zu analysieren versuche, ist sein mereologischer Charakter. Wir sind nämlich im Falle mentaler Selbstrepräsentation mit einer interessanten Variante von *Teil-Ganzes-Relationen* konfrontiert: Ein Teil eines Systems fungiert als interne Abbildung des Systems als Ganzem (und zwar im Interesse des Systems). Ein Teilbereich seiner physischen Aktivität (zum Beispiel in seinem zentralen Nervensystem) steht in einer Ähnlichkeitsrelation zum System als Ganzem.[3] Vieles, was ich im vorangegangenen Kapitel über mentale Repräsentation und analoge Darstellungsformen gesagt habe, gilt *mutatis mutandis* für den Sonderfall, der jetzt unser Thema ist. Unter anderem gilt auch, daß sich das System durch Erzeugung innerer Repräsentate bis hinunter auf die physische Ebene ständig selbst verändert. Für komplexe Selbstrepräsentationssysteme vom biologischen Typ gilt: Selbstmodellierung geht einher mit Selbstveränderung. Diese psychophysische Prozessualität bedeutet für den Gehalt des Selbstrepräsentats, daß er – da die ihn erzeugenden physischen Vorgänge eine zeitliche Dimension besitzen – der Wirklichkeit immer schon hinterherhinkt. Das Selbstrepräsentat kann aus prinzipiellen Gründen niemals ein vollständiges Repräsentat werden, da es Resultat einer Systemveränderung ist, die nicht in „Echtzeit" eingefangen werden kann.[4]

Die Klasse der mentale Selbstrepräsentate erzeugenden physischen Systeme überschneidet sich **drittens** mit einer anderen Klasse von formalen Strukturen, die vornehmlich durch die Mathematik erforscht worden sind, nämlich mit selbstähnlichen Strukturen. Ich habe bereits in Abschnitt 2.2.1 darauf hingewiesen, daß mentale Modelle eine ganz andere Form von

---

delns analysieren: *Theoretisches Abzielen* besteht in der Aktivierung und Bearbeitung von mentalen Modellen (dem Einsatz von abstrakten Organen bei *inneren* Handlungen) und *praktisches Abzielen* in der Aktivierung von „motorischen Simulaten" (die als kausale Antezedentien von *äußeren* Handlungen mit Hilfe der konkreten Effektoren fungieren).

[3] In einem schwächeren und trivialeren Sinne gilt dies auch für das physikalische Universum als Ganzes: Dadurch, daß Systeme in ihm auftauchen, die interne Modelle der Welt erzeugen, wird es zu einem (schwach) selbstähnlichen System – es gewinnt eine neue Qualität hinzu.

[4] Wie aus den Versuchen von Libet 1973, 1979 *et al.* hervorgeht, sind phänomenales Zeiterleben und die faktische zeitliche Strukturierung des Informationsflusses im Gehirn nicht dasselbe. Für mentale Repräsentate im allgemeinen gilt nicht unbedingt, daß ihr Gehalt (abgesehen von seiner Datierung) „der Welt hinterherhinkt". Wenn man nämlich annimmt, daß Zustände der Welt sich manchmal über einen gewissen Zeitraum nicht ändern, dann kann es prinzipiell auch eine Episode des mentalen Repräsentationsvorgangs geben, in der es ein aktives Komplett-Repräsentat dieses Zustandes gibt, das den betreffenden Weltausschnitt einfängt.

Erst Wissen und Intentionalität erzeugen als externe, digitale Repräsentate, nämlich Wissen durch Ähnlichkeit. Für die Frage nach der Selbsterkenntnis (dem Selbst-Wissen) natürlicher Modellgeneratoren bedeutet das: Selbstmodellierende Systeme erzeugen interne Strukturen, die dem System als Ganzem wegen dem Bestehen bestimmter Isomorphismen ähnlich sind. Damit aber ist das System als Ganzes – dadurch, daß es subjektives Bewußtsein instantiiert – zu einer *selbstähnlichen* Struktur geworden. Anders ausgedrückt: Da die mereologische Selbstrepräsentation des Systems (in unserem Fall) durch mentale Modellierung – also durch Analogrepräsentation – erreicht wird, stehen das Ganze und der es intern abbildende Teil in einer Ähnlichkeitsrelation. Diese von einem informationsverarbeitenden System erzeugte Selbstähnlichkeit ist fundamental für ein Verständnis des psychischen Phänomens „Selbstbewußtsein".

Viertens haben von physischen Systemen generierte Selbstrepräsentate eine Eigenschaft, die den Elementen des internen Weltmodells fehlt: Sie können niemals vollkommen *leer* sein. Wenn es überhaupt ein inneres Modell des Selbst gibt, dann muß es nach den Grundannahmen einer naturalistischen Theorie des Geistes auch ein physisches System geben, das es erzeugt hat. Dies ist die naturalistische Variante des cartesianischen *Cogito*: Es kann pathologische oder epistemisch völlig leere Selbstrepräsentate[5] geben – aber die ihnen zugrundeliegende „Existenzannahme" wird niemals falsch sein, weil es irgendein konstruierendes System geben muß. Selbst wenn ich ein Gehirn im Tank, ein Gedanke im Geiste Gottes oder der Traum eines Marsmenschen bin – aus teleofunktionalistischer Perspektive sind Repräsentate immer nur solche *im Kontext eines sie erzeugenden Systems.*

Mentale Modelle von Teilen der Welt dagegen könnten immer auch pure Simulate oder das Ergebnis von Fehlrepräsentationen sein, denn grundsätzlich gibt es für das System keine Garantie dafür, daß es nicht bestimmten seiner inneren Zustände intentionalen Gehalt fälschlicherweise zuschreibt. Selbstmodellierung auf der anderen Seite besitzt einen prinzipiell höheren Gewißheitsgrad und dies ist die moderne, naturalistische Formulierung der cartesianischen Selbstgewißheit des Subjekts. Wenn Descartes im achten Paragraphen der *Zweiten Meditation* noch das Denken als vom Ego untrennbar entdecken konnte, so wird aus heutiger Perspektive das Ego zum *Gedachten*, das untrennbar mit dem *es denkenden* physischen System verknüpft ist. *Dieses* System, zum Beispiel das Gehirn eines biologischen Organismus, ist in Wirklichkeit das denkende Ding. Es erzeugt *cogitationes* in Form mentaler Modelle. Da es aber diese Modelle intern nicht *als* Modelle darstellt, erkennt es sein phänomenales Ich – das mentale Modell einer *res cogitans* – nicht als Produkt innerer Repräsentationstätigkeit, sondern verwechselt es mit sich selbst. Diese Tatsache deutet auf eine grundlegende selbstreferentielle Opazität innerhalb phänomenaler Subjek-

[5] Vgl. Abschnitt 3.2.2.

tivität, auf ein naiv-realistisches Selbstmißverständnis – produziert durch ein selbstmodellierendes physisches System. Ist das Selbstrepräsentat eines Systems nämlich so zuverlässig und schnell aktivierbar, daß es nicht mehr als Konstrukt erlebt werden kann, dann entsteht auch hier wieder die van Gulicksche „semantische Transparenz". Um diese Metapher auf die Spitze zu treiben, könnte man sagen: Als selbstbewußte, phänomenale Subjekte *schauen wir durch uns selbst hindurch* und erkennen uns nicht als komplexe Makro-Repräsentate. Bevor wir tiefer in diese Problematik eindringen, sollten wir wieder einen Blick auf die logische Struktur der verschiedenen Varianten von mentaler Selbstrepräsentation werfen. Wir können uns dabei jedoch kurz fassen, weil wir den wesentlichen Punkten bereits begegnet sind.

3.1.2 Mentale Selbstsimulation: Interne Selbstähnlichkeit und Entwürfe des Selbst

Mentale Selbstrepräsentate sind von Gehirnen erzeugte und benutzte Werkzeuge. Biosysteme setzen diese ein, um schnell und effektiv möglichst viel überlebensrelevante Information über sich selbst zu verarbeiten. Natürlich können auch mentale Selbstrepräsentate unabhängig vom Strom des aktuellen Inputs aktiviert und in mögliche phänomenale Welten eingebettet werden. Gehirne können *kontrafaktische phänomenale Selbste* erzeugen, und diese können als Werkzeuge in absichtlich eingeleiteten kognitiven Operationen eingesetzt werden. Sie können dem System bei der Zukunftsplanung, sowohl in Hinblick auf die Auswertung von Zielzuständen als auch auf die Erzeugung der adäquaten Handlungsmuster behilflich sein. Selbstsimulate treten aber auch als Agenten in inneren Monologen, Phantasien oder Tagträumen auf. In all diesen Fällen sind Selbstsimulate jedoch nur imaginiert, d. h. sie sind Bestandteile einer komplexen mentalen Simulation einer möglichen Welt. Parallel zu diesem mentalen Weltsimulat existiert immer noch ein internes Selbstrepräsentat des Systems *als eben jene mentalen Operationen gerade absichtlich durchführendem*. In solchen Situationen wissen wir als phänomenale Wesen, *daß* wir phantasieren, *daß* wir tagträumen, *daß* wir innerlich mit uns selbst sprechen, *daß* wir an einem inneren Lebensentwurf arbeiten, daß wir uns bemühen, ein Nagelsches *objective self* in uns entstehen zu lassen und so weiter. Das muß nicht so sein – es gibt Zustände, in denen das Selbst ein reines Simulat ist und folglich gibt es in diesen inneren Situationen auch keine Klarheit über den gegenwärtigen Zustand. Das ist zum Beispiel bei Träumen und manchen Geisteskrankheiten der Fall.

Mentale Selbstsimulation: $\mathbf{M}_{\text{S-Sim}}$ (S,X,S)
- X simuliert S *für* S.
- X ist ein mentales Selbstsimulat; d. h. es ist das mentale Repräsentat eines kontrafaktischen Zustands des Systems als Ganzem.
- das teleologische Zusatzkriterium muß nicht erfüllt sein, denn es gibt *afunktionale* Selbstsimulate.
- $\mathbf{M}_{\text{S-Sim}}$ ist antisymmetrisch.

Träume und komplexe Halluzinationen sind solche Zustände, in denen wir uns *verlieren* können, weil das System – das wir sind – sich in ihnen nur noch über ein „kontrafaktisches" Selbstmodell intern gegeben ist. Daß diese kritische Distanz zu unseren eigenen inneren Bildern von uns selbst allein durch Ereignisse auf der physischen Ebene vollkommen zerstört werden kann, ist die eigentlich demütigende Einsicht, zu der uns das neue, durch die Neuro- und Kognitionswissenschaften gezeichnete Bild vom Menschen zwingt: Es sind – wenn wir von der Systemperspektive wieder zur Subjektperspektive wechseln – eben *nicht* unsere *eigenen* Bilder von uns selbst.[6] Ich werde in Abschnitt 3.2.2 auf jene Zustände eingehen, in denen die Ohnmacht des phänomenalen Subjekts gegenüber dem es konstruierenden System deutlich wird und im Schlußkapitel auf diesen Punkt zurückkommen.

3.1.3 Mentale Selbstpräsentation: Verkörperung

Teilweise kann ein System sich selbst auch in Form mentaler Präsentate, also durch output-entkoppelte, nicht-relationale mentale Zustände intern selbst gegeben sein. Das komplexe Selbstrepräsentat kann – wie andere mentale Modelle auch – einen reinen Präsentationsaspekt besitzen, dessen Grundlage vielleicht die dem jeweiligen Modell zugrundeliegende Signalstärke ist. Beim Menschen gibt es einen Teil des Selbstmodells, der – solange es *überhaupt* phänomenales Bewußtsein gibt[7] – permanent als *Präsen-*

[6] Nach Sigmund Freuds Kategorisierung der narzißtischen Kränkungen durch die Wissenschaft (vgl. Freud 1947[1917]: Band XII) ist dies die *dritte*, nämlich die *psychologische* Demütigung: In Freuds Worten besteht sie in der Tatsache, *„daß das Ich nicht Herr sei in seinem eigenen Haus"* (a.a.O.: 7f). Gerhard Vollmer bietet eine Kategorisierung der darauf folgenden vierten bis neunten Kränkung an, wobei die gegenwärtigen Umwälzungen im Menschenbild durch das Computermodell des Geistes als Nummer sieben erscheinen. Vgl. Vollmer 1992.

[7] Richard Gregory hat unter kontrollierten Versuchsbedingungen mit Hilfe des Anästhetikums Ketanest eine langsame Reise in die Unbewußtheit angetreten und die beim Zerfall des mentalen Modells der Wirklichkeit und des Selbst auftretenden phänomenalen Zustände genau beschrieben. Vgl. Gregory 1986: 198–208, dazu auch 1988: 260ff sowie seine Bemerkungen zum Verhältnis des philosophischen und des empirischen Bewußtseinsbegriffs in Gregory 1988. Ketamin (und Phencyclidin) sind auch in Zusammenhang mit der Flohrschen Bewußtseinstheorie (vgl. Flohr 1991, 1992b und Abschnitt 2.1.4) interessant, da sie als NMDA-Antagonisten an genau dem Subsystem des Gehirns angreifen, das für die Entstehung phänomenaler Bewußtheit verantwortlich gemacht wird.

tat gegeben ist: Die aus der „Eigenwahrnehmung" des Körpers entstehenden Erlebnisqualitäten unseres *inneren* sensorischen Bewußtseins. In normalen Wachzuständen sind wir immer schon, vom ersten Moment des Aufwachens bis zum Einschlafen, *verkörperte* Erlebnissubjekte. Das heißt: Wir sind uns in einer ganz bestimmten Art und Weise, nämlich durch unser Körpergefühl, gegeben. Es gibt eine Klasse von mentalen Präsentaten, die alle Veränderungen unseres Selbst- und unseres Weltmodells in der Weise einer sinnlichen Gewißheit begleitet und intern die pure Präsenz des Körpers signalisiert. Diese Klasse von Präsentaten bestimmt die phänomenale Qualität unseres Leiberlebens. In veränderten Bewußtseinszuständen kann dem psychologischen Subjekt der Zugriff auf dieses Basiselement des mentalen Selbstrepäsentats verschlossen sein, was zu gravierenden Veränderungen in der Struktur phänomenalen Bewußtseins führt.[8] Das mentale Präsentat unseres Körpers darf aber nicht mit dem Körperschema verwechselt werden: Es ist nicht das räumlich-kinästhetische Modell des Körpers, seiner Position und seiner Bewegungen – jenes ist eine komplexe relationale Struktur, die hauptsächlich räumliche Beziehungen und Schwereempfindungen zu einem internen Simulat verknüpft. Es ist vielmehr der *Signalaspekt* jenes Modells, der die aktuelle Gegenwart des Systems als eines schweren und räumlich ausgedehnten Objekts intern anzeigt.

Mentale Selbstpräsentation: $\mathbf{M}_{\text{S-Prä}}$ (S,X,S)

- X präsentiert S intern *für* S.
- X ist ein Selbstpräsentat; ein nicht simulationsfähiger interner Systemzustand, der nur bei konstantem Input aufrechterhalten werden kann. Dieser Input wird unter Standardbedingungen *intern* erzeugt.
- $\mathbf{M}_{\text{S-Prä}}$ ist antisymmetrisch.

Das *Präsentandum* – der unser Körpergefühl durch afferente Signale aus den Propriozeptoren in Muskeln, Sehnen, Bändern, Gelenken und der Haut auslösende Körper – ist trivialerweise immer gegeben, wenn das mentale Präsentat überhaupt entsteht. Durch diese besondere Situation – die Verankerung in *interner* Signalerzeugung – wird das Selbstpräsentat zu einem der stabilsten und konstantesten Elemente der phänomenalen Realität, weil es der zuverlässigste Teil unseres multimodalen inneren Bildes von uns selbst ist. Auch wenn wir nicht denken, keine Gefühle haben und uns nicht bewegen: Die permanente Signalquelle, auf der unser Körpergefühl beruht, stellt ihre Aktivität nicht ein – wir sind uns „immer schon" als leibliche und verkörperte Wesen gegeben. Aus diesem Grund empfinden viele Menschen ihr phänomenales Körper-Selbst als den *gewissesten* Teil ihrer selbst und aus demselben Grunde werden Störungen der zugrundeliegenden neuronalen Funktionen als besonders bedrohlich erlebt.[9]

[8] Vgl. Abschnitt 3.2.2.

[9] Auf einer Störung von Selbstpräsentation beruhende Veränderungen im Selbstmodell komme ich in Abschnitt 3.2.2 zurück.

Das Körpergefühl besitzt beim Menschen die bereits erwähnte „instantane“ Qualität, weil der Vorgang seiner Aktivierung nicht noch einmal phänomenal modelliert wird. In komplexe mentale Simulationen geht es normalerweise nicht ein. Andererseits ist es eng verknüpft mit einer phylogenetisch sehr alten Klasse von mentalen Modellen, deren Auftreten deutlich mit Aktivitäten des limbischen Systems korreliert ist: *Emotionen*. Emotionen kann man als sehr schnelle und stammesgeschichtlich alte (d. h. für die *biologische* Umwelt wahrscheinlich funktional stark optimierte) mentale Modelle der gegenwärtigen Interessenlage des Organismus bezeichnen. In den Worten von Robert Ornstein: *„Emotionen waren hier, bevor wir es waren“*.[10] Wegen dieser engen Verwandschaft können emotionale Simulationen auch spezifische Körpergefühle erzeugen. Allgemein gilt jedoch, daß das mentale Präsentat des Körpers nicht simulationsfähig ist, weil es auf spezifischem Input beruht. Wenn wir uns mit geschlossenen Augen einen Salto vom Siebenmeterbrett vorstellen, löst dies normalerweise nicht die entsprechenden kinästhetischen Qualia auf der phänomenalen Ebene aus. Es ist jedoch denkbar, daß mentale Simulationen eines anderen Typs den andauernden Informationsfluß aus afferenten, propriozeptiven Signalen modulieren und so auch für die qualitative Veränderung unseres erlebten Selbstpräsentats verantwortlich sein können. All dies sind Fragen, die in das Gebiet empirischer Einzeldisziplinen fallen. Die Antworten, die diese Disziplinen uns geben, werden zukünftig zu immer genaueren begrifflichen Unterscheidungen zwingen und uns vielleicht ganz nebenbei auch zu besseren Phänomenologen machen, als wir es je waren.[11]

Wenden wir uns nun jedoch wieder der Frage zu, was für eine mentale Struktur es ist, die die in den vergangenen drei Abschnitten skizzierten Funktionen in sich vereint, und welche Bedeutung diese Struktur für eine philosophische Theorie der Subjektivität haben könnte.

3.2 Was ist ein mentales Selbstmodell?

Ein Selbstmodell ist ein in ein internes Modell der Welt eingebundenes Analogrepräsentat *des es konstruierenden Systems in seiner Umwelt*. Sein Gegenstand ist ein diskretes physikalisches System, und hierbei handelt es

[10] Vgl. Ornstein 1986.

[11] Daß eine materialistische Revolution zu einer *Bereicherung* unseres Innenlebens führen wird, ist eine These, die Paul Churchland immer wieder vertreten hat: *„I suggest, then, that those of us who prize the flux and content of our subjective phenomenological experience need not view the advance of materialist neuroscience with fear and foreboding. Quite the contrary. The genuine arrival of a materialist kinematics and dynamics for psychological states and cognitive processes will constitute not a gloom in which our inner life is suppressed or eclipsed, but rather a dawning, in which its marvellous intricacies are finally **revealed** – most notably, if we apply [it] ourselves, in direct self-conscious introspection.“*(Churchland 1989: 66)

sich genau um dasjenige – künstliche oder natürliche – Repräsentationssystem, welches das Selbstmodell *in sich* erzeugt. Mental sind diejenigen Partitionen des Selbstmodells, die prinzipiell durch Metarepräsentation zu Inhalten von phänomenalem Bewußtsein werden können.[12] Derjenige Teil des mentalen Selbstmodells, der durch Metarepräsentation zum *aktuellen* Gehalt phänomenalen Bewußtseins wird, ist das *phänomenale Selbst* und das, was wir in traditioneller Terminologie als das *Subjekt innerer Erlebnisse* zu bezeichnen pflegen.

Mentale Meta-Selbstrepräsentation: $\mathbf{M}_{\text{M-SRep}}$ (S,X,SM)
- X repräsentiert einen Teil des aktiven Selbstmodells *für* S.
- X ist ein interner Systemzustand.
- X ist das aktuelle *bewußte* Modell des Selbst in und *für* S.
- $\mathbf{M}_{\text{M-SRep}}$ ist antisymmetrisch.

Selbstmodelle sind komplexe Repräsentate, mit denen einige informationsverarbeitende Systeme intern auf sich selbst Bezug nehmen. Das ist jedoch nicht im Sinne einer propositionalistischen Referenztheorie der Bedeutung gemeint: Als Resultate von Simulationsvorgängen liegen sie in analogem Format vor, d. h. sie bilden ihre Objekte über eine relationale Struktur und Isomorphismen höherer Ordnung ab. Sie stehen deshalb in einer mereologischen Ähnlichkeitsrelation zu den Systemen, von denen sie generiert werden. Man kann sagen: Selbstmodelle sind physische Realisierungen einer repräsentationalen Teil-Ganzes-Beziehung. Ihre logische Struktur und die ihrer Teilphänomene habe ich in den letzten Abschnitten kurz skizziert. Ich möchte mich deshalb jetzt zunächst der *Genese* von Selbstmodellen zuwenden.

Selbstmodelle in einem starken Sinne kennen wir bis heute nur von Biosystemen.[13] Selbstmodelle sind abstrakte Organe, die von solchen Systemen ausgebildet wurden und ihnen anscheinend einen Auslesevorteil ermöglicht haben. Man muß also fragen: Was könnte die Funktion eines Selbstmodells für ein lebendes System in einer natürlichen Umwelt gewesen sein? Die ersten, rudimentären Selbstmodelle könnten in Zusammenhang mit der Evolution von motorischen Funktionen aufgetreten sein. Sobald ein lebendes System beginnt, sich zielgerichtet zu bewegen[14], wird es

[12] Auf das Selbstmodell gerichtete Metarepräsentationsvorgänge erzeugen auf der Erlebnisebene ein psychisches Phänomen, das üblicherweise als „Introspektion" bezeichnet wird. Ich habe bereits in Abschnitt 2.1 Introspizierbarkeit als Kriterium für die Zuschreibung des Prädikats „mental" angeboten. Damit meine ich nicht, daß Introspektion eine im epistemologischen Sinne zuverlässige Methode des Erkenntnisgewinns ist. Die interne Metamodellierung mentaler Modelle erzeugt keine Form von Wahrheit in irgendeinem interessanten Sinne (vgl. Abschnitt 2.2.1). Was ich hier tue, ist die Definition innerer repräsentationaler Zustände über eine ihrer dispositionalen Eigenschaften – nämlich als Grundlage und kausales Antezedens bestimmter psychischer Phänomene dienen zu können.

[13] Vgl. jedoch Kiefer 1988.

[14] Valentin Braitenberg zeigt in seinem Buch *„Vehicles"* auf amüsante Weise, wie aus der Kombination sehr einfacher Bewegungsmechanismen für den externen Beobachter eines Systems der Eindruck von Zielgerichtetheit entstehen kann. Vgl. Braitenberg 1984.

für das System auch ein Vorteil sein, ein möglichst gutes inneres Bild seines Körpers in seiner Umgebung zu besitzen. Denn sobald ein Biosystem sich in einer Umwelt wiederfindet, in der *koordinierte* und effektive Körperbewegungen einen Selektionsvorteil bedeuten, wird der Besitz eines internen Modells dieses Körpers für es nützlich sein. Dabei ist es wichtig, daß dieses Modell in einer Relation zum Modell der Welt steht: Ein System, das kein Modell der Welt besitzt oder – parallel zu Wahrnehmung der Umwelt – in isolierten solipsistischen Fantasien schwelgt, wird auch seine *Interaktionen* mit dieser Umwelt nicht auf diese Weise optimieren können. Für eine optimale Bewegungskoordination sind zuallererst die geometrischen Eigenschaften des Organismus von Bedeutung. Das bedeutet: Er muß sich intern als *res extensa* modellieren, als räumlich ausgedehntes physisches Ding, welches sich zusammen mit ähnlichen Objekten in einem durch gewisse physikalische Gesetze beherrschten Raum befindet. Dazu ist eine möglichst exakte Repräsentation der Oberfläche bzw. der physischen Systemgrenze nötig. Für komplexeres Verhalten sind dann eine ganze Reihe weiterer körperlicher Eigenschaften relevant: Die Beschleunigung, das Gewicht, die Solidität, die „Kollisionseigenschaften" des Systems etc. Dies sind letztlich empirische Fragen, ich gehe deshalb an dieser Stelle nicht weiter ins Detail.

Man kann jedoch davon ausgehen, daß eine der ersten kognitiven Leistungen in der Unterscheidung von Fremd- und Eigenreizung bestanden hat, also in einer sehr einfachen dualen Kategorisierung mentaler Präsentate.[15] Sobald es Systeme gab, die zur Erzeugung relationaler Repräsentate in der Lage waren, konnten auch die ersten Modelle von räumlichen Relationen bezüglich solcher Fremd- und Eigenreize entstehen, und damit die ersten geometrischen Selbstmodelle. Solche geometrischen Selbstmodelle haben sich wahrscheinlich zu dem entwickelt, was wir heutzutage aus externer Perspektive als das Körperschema einer Person bezeichnen.

Das Körperschema ist – in unserem eigenen Fall – das repräsentationale *Fundament* des Selbstmodells. Neuere Erkenntnisse bezüglich des Schmerzerlebens in Phantomgliedern (z. B. nach Amputationen) deuten auf die Existenz einer genetisch determinierten Neuromatrix[16] hin, deren Aktivitätsmuster die Grundlage des Körperschemas und des Körpergefühls sein könnte. Der inputunabhängige Teil dieses neuronalen Aktivitätsmusters erzeugt ein kontinuierliches repräsentationales Fundament für das körperliche Selbstmodell und wird dadurch zur Grundlage des „gewisse-

[15] Mit anderen Worten: Zum repräsentationalen Gehalt von Selbstmodellen gehört eine bestimmte Teilmenge der durch das System erfaßbaren kausalen Eigenschaften der Welt, nämlich genau diejenigen kausalen Eigenschaften, die vom System sehr leicht, direkt und zuverlässig verändert werden können.

[16] *„In essence, I postulate that the brain contains a neuromatrix, or network of neurons, that, in addition to responding to sensory stimulation, continuously generates a characteristic pattern of impulses indicating that the body is intact and unequivocally one's own. I call this pattern a neurosignature. If such a matrix operated in the absence of sensory inputs from the periphery of the body, it would create the impression of having a limb even when that limb has been removed."* (Melzack 1992: 93) Vgl. auch Melzack 1989, Melzack *et al.* 1997.

sten" Teils des phänomenalen Selbst – darum ist *unsere* Form von Subjektivität fast immer *leibgebundene* Subjektivität. Alle später entwickelten Formen mentaler Selbstrepräsentate treten unter Standardbedingungen stets in Begleitung des räumlichen Modells des Selbst und vor dem phänomenalen Hintergrund des Verkörpertseins auf. Dabei werden sie immer „unräumlicher": Unsere höchsten kognitiven Operationen zeichnen sich nur noch durch eine zeitlich-sequentielle Geordnetheit aus. Denn abstrakte gedankliche Operationen repräsentieren die meisten Menschen mental nur noch im Modus des *Nacheinander* und nicht mehr im Modus des *Nebeneinander*. Diese Tatsache schafft in unserem eigenen Fall das vorphilosophische Leib-Seele-Problem. Das vorphilosophische Leib-Seele-Problem beruht auf einer intuitiven Dissonanz in unserem Selbsterleben, welche uns das *theoretische* Problem der kausalen Wechselwirkung zwischen mentalen und physischen Zuständen vor aller philosophischen Reflexion auch spüren läßt. Es hat – das wird nun deutlich – seine Wurzeln in einem „technischen" Problem humaner Selbstmodellierung: Wie schafft es ein menschliches Gehirn, selbsterzeugte mentale Modelle ohne räumliche Parameter (zum Beispiel den Gedanken *„Ich bin ein Ding, das denkt"*) in ein Selbstmodell einzubetten, das aus Gründen seiner biologischen Genese auf einem *räumlichen* Modell des Systems beruht? Die cartesianische Unterscheidung zwischen denkenden und ausgedehnten Substanzen ist für uns Menschen genau deshalb nicht unplausibel, weil sie eine Entsprechung in der repräsentationalen Struktur unserer Selbstmodelle besitzt.

Wir sind Systeme, die sich selbst erklären müssen, wie es möglich ist, daß sie abstrakte kognitive Operationen mit nicht-sensorischen Repräsentaten zweiter Ordnung durchführen. Wir tun dies, indem wir ein *metakognitives Selbstmodell* aktivieren: Wir erzeugen ein mentales Modell von uns selbst *als Gedanken und begriffliche Erkenntnisse produzierende Wesen*. Metakognition funktioniert mit einem solchen Modell eines denkenden Selbst, welches mentale Modelle eines ganz bestimmten abstrakten Typs erzeugt. Da diese Partition des Selbstmodells erlebnismäßig nicht direkt mit dem mentalen Bild unseres Körpers verbunden werden kann, weil sie eine andere Funktion *für* das System besitzt (nämlich die Überwachung kognitiver Operationen höherer Ordnung), entstehen eine Reihe von intuitiven Dissonanzen. Da außerdem diese Partition des Selbstmodells als einer denkenden, nicht-räumlichen und außersinnlich gegebenen Instanz so schnell aktiviert wird, daß sie auf der phänomenalen Ebene als vor aller inneren Erfahrung immer schon gegeben erlebt wird, bietet sie sich zu transzendentalphilosophischen Interpretationen verschiedener Art an. Ich habe bereits darauf hingewiesen, wie solche phänomenologischen Fehlschlüsse von der Gehaltskonstanz mentaler Repräsentate auf ihren ontologischen Status entstehen und werde diesen Punkt deshalb hier nicht noch einmal wiederholen. Es gibt aber andere Klassen von mentalen Modellen, die wesentlich leichter mit dem mentalen Modell des Körpers, dem repräsentationalen Fundament des Selbstmodells, verbunden werden können als kognitive Repräsentate im engeren Sinne.

Bei stammesgeschichtlich früher entwickelten inneren Zuständen wie etwa Gefühlen gelingt die Einbindung ins Selbstmodell noch leichter. Ein Schreck fährt uns *in die Glieder*, die Ankündigung einer Gehaltserhöhung läßt unser Herz *höher* schlagen und bei dem Gedanken an unsere politische Führung bekommen wir ein flaues Gefühl *im Magen*. Daß Emotionen *meine* Emotionen sind, ist schon deshalb klar, weil sie eine diffuse leibliche Lokalisierung aufweisen. Gefühle gehen mit lokalen somatischen Erregungszuständen einher (zum Beispiel Adrenalinausschüttungen im Bereich des Solar Plexus) und diese somatischen Rand- bzw. Entstehungsbedingungen werden mental noch einmal mitrepräsentiert. Aus der Perspektive des Erlebnissubjekts haben Gefühle neben ihrer Räumlichkeit noch ein weiteres Merkmal mit dem Körperselbst gemein: Sie sind nicht besonders *plastisch*.

Was heißt das? Das emotionale Selbstmodell[17] – die Gesamtheit aller die Interessenlage des Systems modellierenden inneren Zustände eines bestimmten Typs – ist in seinem Gehalt durch „höhere", absichtlich eingeleitete kognitive Operationen nur schwer zu beeinflussen. Gefühle kommen und gehen, sie sind dem Zugriff des psychologischen Subjekts weitgehend entzogen. *Emotionale Simulationen* sind schwer durchzuführen, genauso schwer wie das intendierte Erzeugen von Körpergefühlen. Emotionen konfrontieren das phänomenale Ich immer wieder mit seiner Biologizität.[18] Denn Emotionen sind nicht beliebig durch externe Stimuli auslösbar, sie benötigen zu ihrer Aktivierung die interne Repräsentation eines Zielzustandes. Biologische Systeme können, wenn sie ihre eigenen Zustände in ihrer Relation zu intern definierten Zielzuständen mental modellieren, mit weitgehend starren und schwer kontrollierbaren mentalen Modellen wie Panik, Eifersucht oder Verliebtsein konfrontiert werden. Das Auftreten dieser Modelle demonstriert ihnen, daß sie durch bestimmte „biologische Imperative" wie Selbsterhaltung oder Vermehrung *gebunden* – das soll heißen: in ihrer funktionalen Architektur determiniert – sind. Die Funktion von Gefühlen *für* einen Organismus ist es, ihm durch Einbettung in sein Selbstmodell ein schnelles und möglichst zuverlässiges Repräsentat seiner gegenwärtigen Interessenlage zu liefern. Das System ist sich damit erstmals intern als ein Wesen gegeben, das in bezug auf seine physischen Interessen in einem bestimmten Ausmaß festgelegt ist. Das geringe Ausmaß

[17] Die Beziehung zwischen Emotionen und dem Selbst untersucht Rosenthal 1983.

[18] Einer der populärsten Einwände gegen die Möglichkeit von Maschinenbewußtsein ist immer der folgende gewesen: *„Aber ein künstliches System wird niemals **Gefühle** haben!"* Die diesem Einwand zugrundeliegende Intuition beruht auf der Beobachtung, daß nicht-biologische Systeme sich (bisher) nicht in einer genuinen Wettbewerbssituation befinden. Eine solche Wettbewerbssituation würde einige ihrer fundamentalen Interessen (Selbsterhaltung, Vermehrung) automatisch festlegen. Deswegen haben künstliche Systeme gegenwärtig höchstens vom Programmierer eingesetzte „Zielvariablen", aber nicht das subjektive *Erlebnis* der Determiniertheit und Rigidität gewisser innerer Zustände, die ihre jeweilige aktuelle Interessenlage und letztlich ihre Gattungsgeschichte widerspiegeln.

der phänomenalen Plastizität seiner Emotionen verschafft ihm eine Ahnung von dem Grad seiner eigenen funktionalen Determiniertheit.

Wir sind jedoch noch wesentlich mehr als Wesen, die einen Körper und biologische Interessen besitzen: Wir sind informationsverarbeitende Systeme, die intern mit nicht-sensorischen Repräsentaten zweiter Ordnung operieren. Wir fassen mentale Modelle von über die Sinnesmodule gegebenen Gegenständen zu Klassen zusammen, repräsentieren Relationen zwischen solchen Klassen und erzeugen mentale Modelle von Propositionen oder Sätzen in öffentlichen Sprachen.[19] Unter Standardbedingungen haben diese höherstufigen kognitiven Leistungen – beim Menschen – eines gemeinsam: Sie werden noch einmal zu Repräsentanda eines internen Abbildungsvorgangs, der sie in Elemente des Selbstmodells verwandelt.

Die als Resultate höherstufiger Kognitionsvorgänge entstehenden inneren Zustände haben eine Reihe spezifischer Charakteristika. In bezug auf ihre kausalen Auslösemechanismen gehen sie nicht mit einer körperlichen Erregung einher, deshalb besitzen sie auch selten eine phänomenale Lokalisierung im Körperschema. Dafür gibt es eine einfache Erklärung: Die höheren kognitiven Vorgängen zugrundeliegenden neuronalen Operationen bilden nur solche Zustände ab, die selbst bereits auf komplizierten Ereignisfolgen *im zentralen Nervensystem* beruhen. Das ZNS selbst ist jedoch frei von jeder Sensorik (so ist zum Beispiel das Gehirn vollkommen schmerzunempfindlich). Darum tragen höherstufige Repräsentate von Gehirnzuständen **erstens** keinen sinnlich-qualitativen Charakter[20] mehr, und **zweitens** besitzen sie nur noch eine sehr schwache erlebnismässige Lokalisierung im Körperschema.[21] Dadurch erhält dieser phylogenetisch junge und für uns Menschen sehr dominante Teil des Selbstmodells aufgrund der in unserem mentalen Modell der Realität verkörperten „Hintergrundannahmen“ automatisch den Charakter von Nicht-Weltlichkeit und von Nicht-Räumlichkeit. Die auf diese Weise entstandene phänomenale Grundstruktur unseres Selbstmodells bildet das Fundament der Intuitionen, die die klassische abendländische Subjektphilosophie zu explizieren versucht hat.

---

[19] Oatley (1988: 383f) argumentiert dafür, daß eine der zentralen Funktionen des Selbstmodells in der Organisation und Strukturierung der Zielhierarchie eines Systems besteht.

[20] Wie fast alle Philosophen wissen, gibt es jedoch sehr konkrete phänomenale Qualitäten, die mit intellektuellen Operationen einhergehen – etwa „Aha-Erlebnisse“. Wenn ein System wie das Gehirn sich nach langen Wegen durch seinen Zustandsraum plötzlich in einem stabilen Zustand wiederfindet, kann diese Zustandsänderung auf der Ebene subjektiven Erlebens durch das Auftreten einer sehr konkreten „intellektuellen Quale“ begleitet werden. Ausgehend von der in Abschnitt 2.1.3 aufgestellten Hypothese **AIT** könnte man sagen: Das entsprechende mentale „Aha!-Präsentat“ hat eine Funktion *für* das System, indem es nämlich ein plötzliches Absinken des inneren energetischen Niveaus signalisiert.

[21] Das Gefühl, alle wichtigen Erkenntnisvorgänge liefen im Kopf ab, ist (bei genauerem Hinsehen) nur eine Extrapolation aus der perspektivischen Organisation unseres visuellen Modells der Welt um ein scheinbar hinter unseren Augen liegendes Zentrum herum.

Das *Cogito* ist nämlich die Beschreibung des phänomenalen Zustandes, der entsteht, wenn solche mentalen Modelle in ein metarepräsentiertes Selbstmodell eingebettet werden, die
– keine räumlichen Parameter besitzen,
– keinem Sinnesmodul direkt zugeordnet werden können,
– extern in Form propositionaler Repräsentate dargestellt[22] werden können.

Für das Gehirn ist die mögliche Zuordnung eines mentalen Modells zu einem Sinnesmodul ein starkes Indiz für die *Externalität* des Repräsentandums, für seine „Weltlichkeit". Die Korrelation eines mentalen Modells (zum Beispiel eines Gefühls) mit einer wahrnehmbaren Veränderung im räumlichen Modell des eigenen Körpers als eines ausgedehnten Objekts gibt ihm die Chance, es mit dem evolutionsgeschichtlichen Fundament des internen Selbstbilds zu verknüpfen. Problematisch ist jedoch für Systeme wie uns selbst die Modellierung gewisser höherstufiger Systemzustände („kognitive Operationen") *als eigene* immer dann, wenn diese keine der beiden oben genannten Eigenschaften besitzen. Diese Situation zwingt das System, sich selbst als ein Ding mit nicht-räumlichen und nicht-weltlichen Eigenschaften intern zu repräsentieren. Man kann sagen: Mit dem Auftreten höherer kognitiver Funktionen war das Gehirn gezwungen, sein Selbstmodell um einen – tendenziell stetig anwachsenden – Bereich zu erweitern, der den Organismus intern als transmundanes, nicht-geometrisches Objekt repräsentiert. Die philosophische Interpretation dieser komplexen neuroinformatischen Entwicklung und ihrer Folgen für die phänomenale Struktur unseres psychischen Lebens erfolgte im abendländischen Kulturkreis durch die Einführung des metaphysischen Subjekts in all seinen Varianten.

Diesen Zusammenhang zwischen *natürlicher* Selbstmodellierung (durch Gehirne) und *theoretischer* Selbstmodellierung (durch philosophische und wissenschaftliche Theorien) näher zu untersuchen, ist an dieser Stelle nicht mein Vorhaben. Vielmehr möchte ich meinen Lesern ein genaueres Gefühl dafür vermitteln, was mit dem neuen Begriff „Selbstmodell" gemeint ist. Wie wir sehen, setzt es sich aus über räumliche Relationen dargestellten Elementen (Körperschema, „Leiberleben", Emotionen) und aus *sequentiell* modellierten Inhalten zusammen (kognitive Operationen, „Gedanken"). Abstrakte kognitive Operationen bringen die Aktivierung solcher mentalen Modelle mit sich, die sich extern leicht in propositionaler Form wiedergeben lassen: Unsere Gedanken lassen sich sprachlich leichter ausdrücken und mitteilen, als unsere Gefühle oder Körperwahrnehmungen. Man kann deshalb solche mentalen Modelle als systeminterne Analogrepräsentate verstehen, mit denen das System *externe* Digitalrepräsentate (etwa Sätze in natürlichen Sprachen) *für sich* darstellt. *Vice versa* kann es manche seiner inneren Analogrepräsentate *für andere Systeme* in äußeren, digitalen Codes

[22] Oder im Sinne von McGinn 1989 durch diese *indiziert* werden können.

präsentieren. Dazu muß es externe physikalische Symbole schaffen, zum Beispiel Schallwellen und Schriftzeichen auf Papier oder Bildschirmen.

Wichtig ist wiederum, an dieser Stelle nicht die unterschiedlichen logischen und epistemologischen Eigenschaften zu übersehen, die dadurch ins Spiel kommen, daß die Repräsentate in unterschiedlichen Formaten vorliegen. Es gibt internes, analoges Selbstwissen, das durch Selbstrepräsentation mit Hilfe von mentalen Modellen erzeugt werden kann. Dieses Wissen beruht auf Ähnlichkeit. Zudem gibt es externes, digital codiertes Wissen, das durch Selbstreferenz mit Hilfe von öffentlichen, sprachlichen Instrumenten zustande kommt. Dieses Wissen beruht auf Wahrheit und Falschheit, auf intersubjektiver Kontrolle und kommt durch *referentielle* Bezugnahme zustande (diesen Punkt untersuche ich im folgenden Abschnitt etwas genauer). Nun ist in unserem Zusammenhang der propositional darstellbaren Elemente des Selbstmodells ein zentraler Punkt, daß Propositionen untereinander in logischen Folgerelationen stehen können, mentale Modelle aber nicht. Mentale Modelle sind in ihrer Aktivierung sehr stark an physische Randbedingungen geknüpft und an die Tatsache, daß sie in repräsentationale Makrostrukturen – wie zum Beispiel mentale Modelle der Welt und des Selbst – eingebettet werden. Für Sätze in öffentlichen Sprachen gilt dies nicht.

Wie könnte ein Gehirn die Folgerelationen zwischen Sätzen intern modellieren und sich so die logische Dimension der Wirklichkeit erschliessen? Eine verwandte Frage kann man auch für abstrakte Automaten aufwerfen: Wie könnte man die formalen Folgerelationen, die von einer Turing-Maschine verarbeitet werden, *physikalisch realisieren?* Man tut es, indem man die formalen Folgerelationen der virtuellen Maschine bei der Implementation durch *zeitliche* Folgerelationen innerhalb eines realen Automaten ersetzt. Wenn auch Gehirne nach diesem Lösungsprinzip verfahren (und es gibt gute Anzeichen dafür, daß dies so ist), dann bilden sie logische Eigenschaften digitaler Repräsentate außerhalb des Systems ab, indem sie diese auf interne Zeitrelationen projizieren: Logisch denken heißt, die richtigen mentalen Modelle in der richtigen *zeitlichen* Reihenfolge zu aktivieren.

Ich hoffe, daß aus dem bisher Gesagten bereits hervorgeht, was es heißt, daß Selbstmodelle relationale Strukturen sind. Sie modellieren zeitliche, räumliche, kausale und logische Relationen anhand des inneren Repertoires, das das jeweilige physische System ihnen zur Verfügung stellt. Selbstmodelle besitzen auch Präsentationsaspekte, speziell über die nicht-relationalen Aspekte des Körperrepräsentats – z.B. über das Schweregefühl.[23] Sie können außerdem Bestandteile komplexer mentaler Simulationen sein. Das bedeutet, daß große Partitionen des Selbstmodells (ohne die *aktuelle* Präsenz der jeweiligen Repräsentanda über die Standard-Kausalketten) aktiviert werden können. Interessant ist hierbei, daß in solchen psychischen Situationen normalerweise sofort ein Meta-Selbstmodell entsteht,

[23] Vgl. hierzu Abschnitt 3.1.3.

welches den Vorgang der Selbstsimulation „beobachtet“, steuert und sich selbst als den *mentalen Agenten* betrachtet. Zumindest ist dies sein Gehalt: Es ist das Selbstmodell eines Systems, das jetzt gerade *absichtlich* eine Selbstsimulation durchführt.[24] Ist ein solches, die kritische Distanz zum eigenen psychischen Zustand schaffendes Meta-Selbstmodell nicht vorhanden, so entstehen Zustände des phänomenologischen Typs „Traum“ oder „komplexe Halluzination“. Ich werde auf solche Zustände im übernächsten Abschnitt zurückkommen und dort auf verschiedene Formen der mentalen Simulation des Selbst hinweisen.

Ob wir es nun mit Selbstmodellierung als Repräsentationsprozeß oder bloß mit Selbstmodellierung als Simulation zu tun haben: Immer wird die Struktur des mentalen Realitätsmodells grundlegend verändert. Mit der Aktivierung eines stabilen Selbstmodells geht nämlich die Instantiierung völlig neuer phänomenaler Eigenschaften durch das System einher. Und zwar kann man die Funktion, die Selbstmodellierung für ein System besitzt, auch als die *Zentrierung des Weltmodells* bezeichnen. Mit anderen Worten: Selbstmodellerzeuger können ein *zentriertes Bewußtsein* instantiieren. Was das heissen soll, werde ich jetzt zu erläutern versuchen.[25]

Mit der Aktivierung eines Selbstmodells innerhalb eines Repräsentationen konstruierenden Systems werden *innere* Erfahrung und Introspektion überhaupt erst möglich. Aus einem Realitätsmodell wird nun ein zentriertes Realitätsmodell und dadurch entsteht überhaupt erst *subjektives* Bewußtsein in der uns bekannten Form. Wenn ein Wirklichkeitsmodell sich schrittweise zu zentrieren beginnt, indem ein mehr oder weniger reiches Submodell des es konstruierenden physischen Systems in es integriert wird, steigt die Zahl der in ihm abbildbaren Sachverhalte drastisch an. Der durch das System geöffnete Repräsentationsraum wird sozusagen um die „Dimensionen“ Ich und Welt, Zukunft, Gegenwart und Vergangenheit erweitert – sein Auflösungsvermögen steigt an. Introspektion wird nun möglich, weil es erstmalig *innere* und *äußere* Zustände gibt. Gleichzeitig entsteht so auf der vorphilosopischen Ebene psychischer Realität zum ersten Mal das Leib-Seele-Problem. Der Raum wird differenziert in „die Welt“ und in „*meinen* Körper“; die Zeit zerfällt in ein „Jetzt“ und in Vergangenheit (alle gespeicherten Weltmodelle des Systems) oder Zukunft (komplexe mentale Simulate mit unterschiedlichen Wahrscheinlichkeiten).

Die Qualität des *„Wie es ist, ein X zu sein“*, der wir im nächsten Kapitel in Zusammenhang mit Thomas Nagels Arbeiten zur Subjektivität mentaler Zustände begegnen werden, tritt überhaupt erst im Kontext eines Selbst-

[24] Ich habe bereits mehrfach darauf hingewiesen, daß mentale Modelle von manchen informationsverarbeitenden Systemen eingesetzte *Werkzeuge* sind. Wenn das stimmt, dann ist der kognitive *Agent*, der das Selbstmodell erzeugt, das *System als Ganzes*. Man könnte auch sagen: Personen erzeugen aus bestimmten Gründen phänomenale Personen, um sich selbst intern darzustellen. Insofern ist das im Text angesprochene Meta-Selbstmodell epistemisch leer oder falsch, wenn es die *phänomenale* Person als den kognitiven Agenten modelliert.

[25] Vgl. auch Metzinger 1989.

modells auf. Die Frage *„Wie ist es, ein nicht-zentrierter Bewußtseinsraum zu sein?"* scheint sinnlos.[26] Wenn *„Wie es ist, ein X zu sein"* gleichbedeutend ist mit *„Wie es ist, das von X konstruierte Selbstmodell (in dem von X konstruierten Weltmodell) zu sein"*, dann ergibt sich unter anderem die folgende Frage: Ist die phänomenale Eigenschaft der Subjektzentriertheit mentaler Zustände logisch verknüpft mit ihrem qualitativen Gehalt? Gehört das Auftreten von Qualia zum *Begriff* des psychologischen Subjekts in einer modernen, naturalistischen Theorie des Geistes?

Ich meine, wir müssen uns vor jeder Form von Humanchauvinismus hüten. Tatsächlich ist es wahr, daß Selbstbewußtsein in unserem eigenen Fall fast ausschließlich mit qualitativem Gehalt einhergeht. Es ist sogar so, daß die allermeisten von uns sich Selbstbewußtsein ohne Emotionen und Körpergefühl, ohne Farben, Klänge und Gerüche nicht einmal vorstellen können.[27] Aus der in den vorangegangenen Abschnitten skizzierten Theorie mentaler Repräsentation geht jedoch eines hervor: Imaginierbarkeit ist nicht dasselbe wie Analytizität. Daß wir uns etwas *vorstellen* können, bedeutet nämlich nur, daß unsere Gehirne die entsprechenden mentalen Simulationen problemlos durchführen können. Da mentale Modelle jedoch keine Syntax besitzen und untereinander nicht in Folgerelationen stehen, kann man aus der Möglichkeit – oder Unmöglichkeit! – gewisser mentaler Simulationen nicht auf *logische* Möglichkeit oder Unmöglichkeit verwandter propositionaler Wahrheiten schließen.[28]

Es ist durchaus denkbar, daß im menschlichen Wachbewußtsein qualitativer Gehalt und Subjektivität nur deshalb immer koinstantiiert sind, weil die neurobiologischen Randbedingungen für mentale Präsentation und mentale Selbstmodellierung identisch oder in ihrem Auftreten eng verknüpft sind. Aus dieser Tatsache dürfen keine voreiligen Schlüsse für eine generelle, philosophische Theorie des Geistes gezogen werden. In jedem Fall ist es richtig, zu sagen, daß durch interne Selbstmodellierung eine faszinierende neue Eigenschaft phänomenalen Bewußtseins erzeugt wird, die uns ein philosophisches Rätsel aufgibt. Eine befriedigende naturalistische Theorie des Geistes wird uns die Lösung dieses Rätsels bieten müssen – oder sollte uns zumindest erklären, *warum* wir es immer falsch verstanden haben. Die problematische neue Eigenschaft habe ich als Zentriertheit

[26] Sicher *gibt* es solche Klassen von Realitätsmodellen, etwa bei Tieren, Kleinkindern oder im Falle mystischer Bewußtseinszustände. Für Systeme, die solche nicht-zentrierten Repräsentationsräume in sich öffnen, sind bestimmte theoretische Fragestellungen jedoch *a priori* unzugänglich. Ein System ohne Selbstmodell kann Thomas Nagels philosophisches Problem nicht erfassen.

[27] Die Frage, die sich hier stellt, ist die nach der Möglichkeit qualia-freier Realitätsmodelle. Wenn ich recht habe und es prinzipiell auch komplexe mentale Modelle der Realität *ohne Präsentate* (vgl. Abschnitt 2.1.3) geben könnte, dann ist auch so etwas wie ein nicht-sinnliches, nicht-leibgebundenes phänomenales Bewußtsein logisch möglich. Vielleicht könnten sich Großrechner oder Mathematiker im Mathematikerhimmel einer solchen psychischen Lebensform ohne qualitativen Gehalt erfreuen – oder auch nicht.

[28] Dieser Fehler unterläuft sogar Saul Kripke in seinem Argument gegen die Identitätstheorie. Vgl. Kripke 1971, 1972.

bezeichnet, und zwar um anzudeuten, daß sie dem erlebten Modell der Welt einen Mittelpunkt verleiht. Das Rätsel besteht in der bereits im ersten Kapitel formulierten Frage: *Wie können in einer objektiv mittelpunktlosen Welt derartige phänomenale Zentren entstehen?* Im nächsten Kapitel werden wir dieser Frage erneut begegnen.

Selbstmodelle sind physisch realisierte Datenstrukturen, die von informationsverarbeitenden Systemen intern aufgebaut werden. Dadurch, daß sie in eine umfassendere interne Datenstruktur eingebettet werden, entsteht bezüglich des so veränderten Gesamtmodells der Wirklichkeit eine neue Eigenschaft: Es wird nun zu einem *perspektivischen* Modell der Wirklichkeit. Diese „Perspektivität" ist aufs engste verwandt mit dem, was ich Zentriertheit genannt habe. Ich glaube andererseits nicht, daß es „perspektivische Fakten" im Nagelschen Sinne gibt und werde später gegen diesen ungerechtfertigten Schluß argumentieren. „Perspektivität" ist aber eine sehr fruchtbare phänomenologische Metapher, weil sie genau den essentiellen und doch relationalen Aspekt psychologischer Subjektivität einfängt, der eine so große Schwierigkeit für post-metaphysische Theorien des Selbstbewußtseins darstellt. Selbstmodelle sind möglicherweise diejenigen empirisch verankerbaren theoretischen Entitäten unterhalb der personalen Beschreibungsebene, die wir benötigen, um das theoretische Rätsel zu lösen. Denn Selbstmodelle treten nicht isoliert auf, sie verändern repräsentationale Gesamtzustände. Das tun sie, indem sie ihrerseits einen Teil der vom System konstruierten mentalen Repräsentate verbinden und so auf ein psychologisches Subjekt – *dessen* Zustände sie nun werden – zentrieren. Man kann sagen: Mentale Selbstmodellierung durch Systeme unseres eigenen Typs ist genau der Sonderfall von repräsentationaler Objektbildung, der zugleich auch zu einer *Subjektbildung* führt.

Fassen wir noch einmal zusammen. Selbstmodelle sind komplexe Datenstrukturen, die innerhalb größerer Datenstrukturen auftreten und diese verändern. Ihre Funktion besteht unter anderem darin, daß sie solche dynamischen, sich ständig verändernden Datenstrukturen zentrieren. Diese Zentriertheit des repräsentationalen Gesamtzustandes bringt neue psychische Phänomene mit sich. Zum Beispiel entsteht ein Erlebnisbrennpunkt, und dadurch erhalten manche internen Repräsentate eine neue Qualität: *Meinigkeit*. Beim Menschen besteht das phylogenetische Fundament des Selbstmodells wahrscheinlich im Körperschema, also in einem räumlichen Modell des es konstruierenden Biosystems. Später kommen mentale Modelle von abstrakteren Eigenschaften des Organismus – zum Beispiel seiner Interessenlage – hinzu, welche das Selbstmodell schrittweise anreichern und die mehr oder weniger plastisch sind. Schließlich hat sich beim Menschen ein kognitiv-intellektuelles Selbstmodell entwickelt. Dessen Bestandteile sind durch höherstufige Operationen generierte mentale Modelle, die keine räumliche Codierung mehr besitzen und auch nicht mehr einzelnen sensorischen Systemen zugeordnet werden können. Selbstmodelle sind durch funktionale Zustände erzeugte interne Datenstrukturen, die eine Funktion *für* das System besitzen. Bei biologischen Systemen besitzen sie

eine biologische Funktion, sie können als Instrumente oder Waffen interpretiert werden. Sie ermöglichen es einem System, eine relevante Teilmenge seiner Eigenschaften intern zu repräsentieren und dadurch zu überwachen. Da Selbstmodelle in unserem eigenen Fall innere Analogrepräsentate – wenn man so will: komplexe „Bilder“[29] in mehreren Modalitäten – sind, referieren sie nicht auf die sie erzeugenden Systeme, sondern simulieren sie. Was das bedeutet, werde ich im folgenden Abschnitt untersuchen.

3.2.1 Selbstsimulation und Selbstreferenz

Die Erzeugung eines stabilen Selbstmodells ist die Grundlage von Selbstbewußtsein und der zusammen mit ihm entstehenden psychologischen Eigenschaften. Inwiefern jedoch wird durch mentale Selbstmodellierung *Wissen* geschaffen, Wissen um das eigene Selbst? Im zweiten Kapitel habe ich mentale Modelle als von manchen Systemen gebildete und zu gewissen Zwecken eingesetzte abstrakte Instrumente der Wissensgewinnung beschrieben. Dabei habe ich auch auf ein mögliches, empirisch-begriffliches Projekt hingewiesen, welches in der Entwicklung einer *Typologie* mentaler Modelle bestünde. Ein solches Unternehmen würde vor dem Hintergrund einer repräsentationalen Theorie des Geistes eine möglichst genaue Klassifikation von Systemen anhand der von ihnen eingesetzten internen Repräsentate zu entwickeln versuchen. Ein wichtiger Bestandteil müßte hier die Analyse der Formate sein, in denen Repräsentate innerhalb eines Systems vorliegen. Die vom System gewählte Form der Darstellung hat nämlich entscheidenden Einfluß auf psychologische „Makro-Eigenschaften“ desselben. Sie hat aber auch einen Einfluß auf die epistemischen Resultate der kognitiven Operationen, die das System mit Hilfe solcher innerer Werkzeuge der Informationsverarbeitung durchführen kann.

Selbstmodelle sind innere Werkzeuge der Informationsverarbeitung. Wie alle mentalen Modelle haben sie die Aufgabe, die relationale Struktur ihrer Gegenstände so vollständig wie möglich darzustellen. Sie zeichnen sich aber dadurch aus, daß das von ihnen modellierte Objekt eben derjenige kognitive Agent ist, *für* den sie eine wichtige Funktion besitzen, und dadurch, daß dieses Objekt ihnen sowohl über externe als auch über interne Signalquellen gegeben ist: Wir können uns im Spiegel betrachten und verstehen sofort, daß unser visuelles Gegenüber identisch ist mit derjenigen Person, welche uns gleichzeitig phänomenal über das Körpergefühl präsentiert wird.[30] Wie alle mentalen Modelle sind auch Selbstmodelle Analogrepräsentate. Als solche zielen sie auf Ähnlichkeit ab – nicht auf Wahrheit. Selbstmodelle sind nämlich nicht Beschreibungen, sondern innere Bilder.

[29] Die Bild/Wort-Unterscheidung entstammt selbst dem öffentlichen Bereich, *innerhalb* des Gehirns und des von ihm geöffneten phänomenalen Raums muß sie natürlich nicht auftauchen. Deshalb ist mein Gebrauch von „Bild“ nur metaphorisch; vgl. Scholz 1991.

[30] Wir fusionieren also ein neues, über die visuelle Modalität aktiviertes mentales Modell von uns selbst mit einem bereits bestehenden. Dem subjektiven Erlebnis des Wiedererkennens entspricht die erfolgreiche *Einbettung* des visuellen Modells unseres Körpers im Spiegel in das bereits aktive Selbstmodell. Vgl. dazu Metzinger 1993b und Abschnitt 5.2.

Von den Bildern an den Wänden unserer Wohnungen unterscheiden sich die von unseren Gehirnen gezeichneten Bilder dadurch, daß sie mehrere Modalitäten umfassen und von dem sie erzeugenden Künstler in Abständen von wenigen Millisekunden ständig verbessert und revidiert werden. Sie sind auch dadurch gekennzeichnet, daß sie in komplexe mentale Simulationen eingehen können und daß sie – in unserem eigenen Fall – in regelmäßigen Zeitabständen vollkommen verschwinden.

Selbstmodelle ermöglichen einem System eine ganz spezifische Form von Selbsterkenntnis. Sie verhelfen dem System zu *analogem* Wissen über sich selbst. Ein solches analoges Selbstwissen beruht – *mutatis mutandis* gelten hier die meisten meiner in Abschnitt 2.2.1 über den Unterschied zwischen analogem und digitalem Wissen gemachten Bemerkungen – auf einem epistemischen Vorgang des Typs „Simulation durch interne Modelle". Dieser Vorgang ist grundverschieden von demjenigen, bei dem Sprecher natürlicher Sprachen Selbsterkenntnis in Form von wahren Sätzen über sich selbst zu erreichen versuchen.

Ein Selbstmodell ist ein gutes mentales Modell des es konstruierenden Systems, wenn es eine hohe Funktionalität für dieses System besitzt. Für biologische Organismen besteht diese Funktionalität in Überlebenssicherung und hohem Fortpflanzungserfolg. Von biologischen Systemen wie uns selbst entworfene künstliche Selbstmodellerzeuger könnten sich – da sie einer anders definierten Umwelt entstammen – mit ganz anderen Zielen selbst modellieren. Die Funktionalität eines Selbstmodells ist jedoch nicht gleichbedeutend mit epistemischem Reichtum[31], sie manifestiert sich nämlich in optimaler Verhaltenssteuerung relativ zu einer gegebenen Umwelt. Weil die von menschlichen Gehirnen erzeugten Selbstmodelle einem Millionen Jahre dauernden „kognitiven Wettrüsten"[32] auf unserem Planeten entstammen, das relativ zu der spezifischen biologischen Umwelt des Menschen auf diesem Planeten zu einer Optimierung von Funktionalität führte, kann man nicht davon ausgehen, daß sie grundsätzlich einen größeren epistemischen Gehalt besitzen als die Ergebnisse anderer Selbstrepräsentationsprozesse – etwa Beschreibungen unserer selbst durch wissenschaftliche Theorien. Wenn sie jedoch einen hohen Wissensgehalt besitzen, dann dadurch, daß sie einen optimalen Grad von Ähnlichkeit erreichen. Wir erinnern uns: Ein gutes internes Simulat ist im Normalfall durch seine problemlose Einbettbarkeit in das aktuelle Weltmodell charakterisiert. Auch für Selbstmodelle gilt, daß das System die Tatsache, daß sie sich problemlos in die relationale Gesamtstruktur des internen Realitätsmodells einfügen lassen, als Indiz für eine weitgehende Isomorphie zwischen innerer Struktur und Repräsentandum wertet.

---

[31] In den Worten von Marcel Kinsbourne: *„If the concept of self evolved, it did so on account of adaptive advantage, not because it reflects some objective truth. The concept of self reifies the organizing activity of a cybernetic device that incorporates its history (‚experience') into the basis for its actions. It is the construct around which are organized impressions and intentions that reach awareness."* (Kinsbourne 1988: 249)

[32] Diese Metapher stammt von Andy Clark. Vgl. Clark 1989: 62.

Ein Selbstmodell verkörpert also das Wissen, das ein System intern über sich selbst gewonnen hat, in einer nicht-propositionalen Form. Dieses Wissen über sich selbst muß nicht mit der Koinstantiierung von Selbstbewußtsein einhergehen, also in Form von mentalen Modellen gegeben sein, die noch einmal durch eine Metamodellierungsfunktion erfaßt werden. In Wirklichkeit ist nur ein geringer Teil des Wissens, das Menschen über sich selbst besitzen, ihnen auch als *aktueller* Inhalt ihres phänomenalen Selbstbewußtseins gegeben. Ein mit traumhafter Sicherheit über den vom Vollmond beschienenen Dachfirst schreitender Schlafwandler ist ein Beispiel für ein System, das ein komplexes, stabiles und funktional aktives Selbstmodell erzeugt, ohne daß in ihm ein phänomenales Bewußtsein entsteht. Wie alle mentalen Modelle besitzen auch Selbstmodelle keine Variablen, keine logische Form und keine Grammatik. Das durch sie erzeugte Wissen ist ein nicht-diskursives Wissen. Wenn das richtig ist, dann stellt das erlebnismäßig nicht hintergehbare jeweilige Modell des Selbst eine innere Form von Wissen dar, die sprachlich nicht oder zumindest nur teilweise und inadäquat ausgedrückt werden kann. Subjektivität läßt sich nicht in den Raum intersubjektiver Diskurse transportieren, weil dieser Raum durch externe, digitale Repräsentate geöffnet wird. Der perspektivische Innenraum des psychologischen Subjekts dagegen wird durch einen völlig anderen Typ von Informationsverarbeitung und Repräsentation erzeugt.

Bereits die Tatsache, daß Sie dieses Buch lesen und verstehen können, zeigt jedoch, daß es mit dieser einfachen Unterscheidung nicht getan ist. Die Situation ist wesentlich komplizierter. Wir wissen bereits, daß Selbstmodelle biologische Instrumente sind, die zu bestimmten Zwecken eingesetzt werden. Auch die Bildung von Gesellschaften war eine erfolgreiche evolutionäre Strategie, setzte aber bei den Mitgliedern solcher Gesellschaften einen bestimmten Typ von interner Selbstbezugnahme voraus. Sie mußten sich mental nun auch bezüglich der für Interaktionen mit der *sozialen* Umwelt wichtigen Eigenschaften ihrer selbst modellieren. In unserem eigenen Fall sind Selbstmodelle deshalb immer auch die Modelle von Sprechern öffentlicher Sprachen, von moralischen Agenten und von *Personen*, die externe, digitale Codes zu vielfältigen Zwecken benutzen. Unsere Gehirne müssen uns unter anderem *als soziale Wesen* mental modellieren[33], als Elemente größerer Systeme. Diese größeren Systeme werden aus Wesen gebildet, von denen vernünftigerweise angenommen werden darf, daß die meisten von ihnen ihrerseits ein perspektivisches, phänomenales Bewußtsein entwickelt haben. Die komplexen Wechselwirkungen zwischen externer und interner Selbstrepräsentation durch verschiedene Mechanismen lasse ich hier außer acht, da sie nur den Gehalt von Selbstmodellen

[33] *„The model of self is not an abstract representation of certain properties of the cognitive system, but a model made up from our experience with the physical and social world. According to this idea, self is a relatively concrete distillation and internalization of our perceptions of physical effects of our actions and of the social impressions of us that people have verbally and non-verbally conveyed back to us."* (Oatley 1988: 378f.)

bestimmen – nicht aber die grundlegende Perspektivität, die diese einem repräsentationalen Gesamtzustand verleihen.

Wenn Selbstmodellierung eine spezielle Form von Wissensgewinnung außerhalb von Sätzen – ohne Wahrheit und Referenz – ist, dann muß man einen Blick auf die Beziehung werfen, die zwischen dem kognitiven Agenten und dem fraglichen Repräsentat besteht. In unserem eigenen Fall sind die kognitiven Agenten, die durch spezielle Formen inneren Handelns nicht-diskursives Wissen über sich selbst erwerben, Personen.[34] Nehmen wir einmal an, daß dies das Folgende bedeutet: Sie sind durch öffentliche Eigenschaften identifizierbare, logisch primitive Entitäten, welche sowohl physische als auch psychologische Eigenschaften besitzen.[35] Diese Entitäten konstruieren mentale Modelle ihrer selbst, die zu Inhalten phänomenalen Bewußtseins werden können und dadurch neue psychologische Eigenschaften erzeugen. Denjenigen Teil des mentalen Selbstmodells, der zum Inhalt bewußter Erlebnisse wird und den ein System generiert, dem wir den Status einer Person zubilligen, werde ich ab jetzt auch als die *phänomenale* Person bezeichnen. Damit ist die Beziehung zwischen kognitivem Agent und Selbstrepräsentat die Beziehung zwischen Person und phänomenaler Person.

Manche Personen – zum Beispiel menschliche Wesen in nicht-pathologischen Wachzuständen – entwickeln durch interne Informationsverarbeitung phänomenale Personen. Phänomenale Personen entstehen durch Erzeugung komplexer Analogrepräsentate, mit deren Hilfe Personen sich selbst mental modellieren. Phänomenale Personen sind die Selbstmodelle von Personen. Vor diesem begrifflichen Hintergrund kann man nun eine Reihe von klassischen Fragen der Philosophie des Geistes als Fragen bezüglich der Relation zwischen Person und phänomenaler Person formulieren. Zum Beispiel könnte man nach der kausalen Relation zwischen „äußerer" und „innerer" Person fragen und vielleicht die folgende Antwort geben: Es gibt keine detailliert beschreibbaren Kausalketten zwischen Person und phänomenaler Person, weil die jeweiligen Beschreibungsebenen inkommensurabel sind. Man kann aber eine Supervenienz der phänomenalen Person bezüglich der sie konstruierenden Person behaupten, indem man sagt, daß **a)** alle Eigenschaften der phänomenalen Person feststehen, wenn alle physikalischen Eigenschaften der Person feststehen, und **b)** wenn zwei phänomenale Personen sich unterscheiden, sich immer auch die sie benutzenden physischen Personen bezüglich ihrer Eigenschaften unterscheiden müssen.[36]

---

[34] Vgl. hierzu auch Pollock 1989.

[35] Vgl. hierzu die in Siep 1983 zusammengestellten Texte, sowie Dennett 1981, Schütt 1981, Strawson 1972. Eine gute Literaturzusammenstellung findet sich in Bieri 1981: 347ff.

[36] Man würde dann davon ausgehen, daß der phänomenale Gehalt des Selbstmodells nicht auf physikalische Eigenschaften des Systems reduziert werden kann, aber trotzdem vollständig von unten determiniert ist. Dies käme dem Versuch gleich, eine nicht-reduktive materialistische Theorie der Subjektivität zu entwerfen. Vgl. jedoch neuere Zweifel an der Stabilität einer solchen Position zwischen den Alternativen eines psychophysischen Dualismus und einer direkten eliminativistischen Strategie in Kim 1989.

Die in unserem Zusammenhang interessante Frage lautet: Ist die Relation zwischen Person und phänomenaler Person eine *epistemische* Beziehung? Wir haben es hier mit einem Sonderfall zu tun, bei dem eine Person sich mental selbst als Person repräsentiert. Das bedeutet: Sie simuliert sich intern als eine Entität mit physischen und psychologischen Eigenschaften, die gewissen philosophischen Theorien zufolge durch einen logisch nicht weiter analysierbaren Begriff – nämlich „Person" – erfaßt wird. Sie tut dies, indem sie eine interne Datenstruktur erzeugt, die neurobiologisch realisiert ist. Die phänomenale Person ist also ein neurobiologisch realisiertes Simulat, das in einer Ähnlichkeitsrelation zu dem physischen System steht, durch das es generiert wird. Während derjenigen Zeiträume, in denen eine Person eine phänomenale Person in sich erzeugt, ist sie demnach auch eine selbstähnliche physische Struktur.

Nun kann man fragen, ob diese dem psychischen Phänomen des Selbstbewußtseins zugrundeliegende interne Selbstähnlichkeit eines informationsverarbeitenden Systems bzw. einer Person als eine Form von Wissen außerhalb von Sätzen gelten darf. Entscheidet man sich für einen engen Wissensbegriff, der öffentliche, propositionale Repräsentate, Wahrheit und Referenz voraussetzt, dann folgt daraus die Opazität von Subjektivität. Was soll das heißen? Es bedeutet, daß unser inneres Erleben der eigenen Zustände prinzipiell nicht als Wissen gelten kann: Unser Selbstbewußtsein ist epistemisch dunkel, die cartesianische Selbsttransparenz des Bewußtseins existiert nicht. Es gibt nur *subjektives Erleben*, aber keine *Selbstgewißheit des Subjekts.* Vielen Philosophen, die an eine radikale Biologizität aller psychologischen Eigenschaften glauben, wird diese Interpretation – auch angesichts der genetischen Erklärung des phänomenalen Subjekts als eines abstrakten Organs, das im Verlauf einer biologischen Wettbewerbssituation entstanden ist – zusagen.

Man kann den Spieß jedoch auch umdrehen und erwidern: Um so schlechter für den philosophischen Wissensbegriff, umso schlechter für die derzeitige Erkenntnistheorie. Die Leistungen, die Organismen durch Selbstmodellierung erbringen können, speziell die Fortschritte des Menschen in bezug auf seine Interaktion mit der Welt und seinen Artgenossen, die ihm durch ein reiches, stabiles Selbstmodell möglich wurden, sind gegenüber früheren Entwicklungsstufen so enorm, daß man sie überhaupt nur verstehen kann, wenn man sie als einen Zuwachs von *Wissen* interpretiert. Aus diesem Grunde muß man die Entwicklung einer Epistemologie natürlicher Repräsentationssysteme fordern.

Daß die Natur der Beziehung zwischen Person und Selbstbewußtsein ein Problem ist, tritt erst auf der Ebene theoretischer Reflexion zutage. Unser durch mentale Modellierung erzeugtes phänomenales Bewußtsein kennt keine kritische Distanz zu sich selbst, weil sich unser Selbstmodell durch die bereits erwähnte van Gulicksche semantische Transparenz[37] auszeich-

[37] Man darf den van Gulickschen Begriff der „semantischen Transparenz" nicht mit dem von Clark verwechseln. Clark möchte mit ihm klassische kognitivistische Systeme im Sinne

net: Das Selbstmodell wird so zuverlässig und schnell aktiviert, daß es mental nicht *als Konstrukt* repräsentiert wird – zumindest gilt dies für seine phylogenetisch alten Elemente. Man kann deshalb sagen, daß wir erlebnismäßig in einem naiv-realistischen Selbstmißverständnis gefangen sind, da wir den Unterschied zwischen Person und phänomenaler Person *subjektiv* nicht machen können. Dieses naiv-realistische Selbstmißverständnis ist ein zentrales Charakteristikum *unserer* Form von Selbstbewußtsein und wird durch die speziellen Mechanismen der Informationsverarbeitung verursacht, die *dieses* Bewußtsein hervorbringen. Es könnte natürlich andere Formen der Selbstrepräsentation geben, die unter Umständen auch andere psychologische Eigenschaften mit sich brächten.

Analytische Philosophen haben sich dem Problem der Subjektivität oft genähert, indem sie die Logik des Wörtchens „Ich" akribischen Untersuchungen unterzogen. *Phänomenale* Subjektivität entsteht jedoch durch interne Analogrepräsentation und nicht durch öffentliche Selbstbezugnahme. Auf der anderen Seite bedienen sich zumindest menschliche Organismen auch externer Repräsentationssysteme, um Wissen über sich selbst zu erwerben oder in Kommunikationssituationen auf sich selbst bezugzunehmen. Externe Repräsentationsprozesse – zum Beispiel die sprachliche Selbstzuschreibung psychologischer Eigenschaften *(„Ich bin jetzt gerade etwas verwirrt")* – beeinflussen den psychischen Zustand von menschlichen Personen ohne Zweifel sehr stark, aber sie tun dies nur indirekt. Nach unserem derzeitigen Kenntnisstand ist es plausibel, anzunehmen, daß die Bewußtseinszustände und psychologischen Eigenschaften von Menschen sehr eng mit den internen, repräsentationalen Gesamtzuständen ihrer Gehirne korreliert sind.[38] Die Entstehung von repräsentationalen Gesamtzuständen folgt jedoch, wie wir gesehen haben, völlig anderen Gesetzmäßigkeiten. Externe Symbolketten beinflussen unsere subjektiven Zustände nur, indem sie durch unsere Gehirne *in Form von mentalen Modellen*, d. h. in analogem Format intern abgebildet werden. Es mag sein, daß wir manche höherstufigen kognitiven Operationen überhaupt erst erlernt haben, weil wir sie zuerst mit Hilfe externer, physischer Symbolsysteme (Schallwellen, Schriftzeichen) erbracht haben und dann die sequentielle Abfolge und Manipulation externer Repräsentate intern durch mentale Modelle

---

von Fodor und Pylyshyn bezeichnen, die eine interne Syntax besitzen und eine regelgeleitete Transformation von Symbolen durchführen, welche auf *formalen* Eigenschaften mentaler Repräsentate beruht. Die Einführung des Terms „semantisch transparentes System" (STS) geschieht hier – in Anlehnung an Arbeiten von Smolensky; vgl. Smolensky 1987, 1988 – mit der Zielsetzung der *Abgrenzung* solcher Systeme gegenüber konnektionistischen Systemen, die repräsentationalen Gehalt intern durch „verborgene Einheiten" darstellen. Vgl. Clark 1989: 17–21, 111–120.

[38] Die Korrelationsthese kann sowohl als Prämisse für Identitätstheorien als auch für dualistische Interpretationen der psychophysischen Relation fungieren. Dies gilt natürlich letztlich auch für die Beziehung zwischen phänomenalen und repräsentationalen Zuständen: Aus ihrer Korreliertheit folgt noch nicht ihre Identität. Empirisch plausibel ist dagegen die These, daß phänomenale Eigenschaften durch bestimmte Kombinationen und höherstufige Verschachtelungen repräsentationaler Zustände instantiiert werden.

„imitiert", bzw. mental *simuliert* haben. Die dieser mentalen Modellierung zugrundeliegenden neuronalen Prozesse dürften ihrerseits subsymbolischer Natur sein – das heißt: auf massiv paralleler Informationsverarbeitung beruhen.[39]

Was wir verstehen müssen, ist die Interaktion zwischen Selbstreferenz und Selbstmodellierung, die Art und Weise, in der die *Logik* des Selbstbewußtseins die *Psychologie* des Selbstbewußtseins beeinflußt – und *vice versa*. Man kann das Rätsel des Selbstbewußtseins nicht auf die Frage einengen, ob es einen referentiellen Gebrauch von „Ich" durch das Subjekt gibt. Andererseits erklärt uns eine naturalistische Theorie der Selbstmodellierung nur das Zustandekommen erlebten Selbstbewußtseins, schlägt aber keine Brücke zur sozialen Dimension von Subjektivität. Diese Dimension öffnet sich durch externe Systeme der Repräsentation und Selbstbeschreibung. Externe Repräsentationssysteme entstehen in Gesellschaften und spiegeln die Geschichte dieser Gesellschaften wider.[40] Das Interface zwischen der inneren Biographie von Individuen, die ihr je eigenes psychisches Drama durchleben, und der äußeren Geschichte von durch die öffentlichen Eigenschaften von Personen gebildeten sozialen Systemen entsteht aber durch genau diese Wechselwirkung zwischen sprachlicher Selbstreferenz und interner Selbstmodellierung. Deswegen möchte ich abschließend noch einen kurzen Blick auf das Problem der Selbstreferenz werfen.

Worauf bezieht sich der Subjektgebrauch von Pronomina der ersten Person? Wenn ich sage *„Ich bin gerade etwas verwirrt"*, worauf bezieht sich dann diese Selbstzuschreibung einer psychologischen Eigenschaft? Befinde ich mich nicht in einer Kommunikationssituation, sondern handelt es sich um einen Fall *monologisierender Selbstzuschreibung*, dann nehme ich damit Bezug auf ein subjektives Erlebnis. Subjektive Erlebnisse entstehen, wie wir gesehen haben, dadurch, daß mentale Modelle in ein Selbstmodell eingebettet werden. Wenn ich zu mir selbst sage *„Ich bin gerade etwas verwirrt"* beziehe ich mich also durch einen Sprechakt auf den Gehalt meines internen Selbstmodells, der Grundlage meines subjektiven Erlebnisses. Es gibt demnach ein von meinem Gehirn konstruiertes internes Repräsentat meiner Verwirrtheit, das in mein aktuelles Selbstmodell eingebettet wurde. Auf den durch dieses mentale Modell meiner Verwirrtheit erzeugten psychischen Zustand bzw. auf den durch die Veränderung des Selbstmodells entstandenen veränderten phänomenalen Gehalt meines Selbstbewußtseins beziehe ich mich mit Hilfe des durch meinen Sprechapparat erzeugten propositionalen Repräsentats. Dieses propositionale Repräsentat wird physikalisch (zum Beispiel durch Schallwellen) außerhalb des Systems realisiert, das ich bin. Das bedeutet: Ein externes Selbstrepräsentat (in propositionalem Format) wird eingesetzt, um auf ein internes

[39] Diese Spekulation äußert auch Clark in seiner Analyse einiger Arbeiten von Rumelhart und Smolensky. Vgl. Clark 1989: 133f, 1987; Rumelhart, Smolensky *et al.* 1986: 40-44, 46; Goschke/Koppelberg 1990: 15f.

[40] Vgl. hierzu Oatley 1988.

Selbstrepräsentat bezugzunehmen. (Achtung! Wenn ich den Satz *„Ich bin gerade etwas verwirrt“* nur *denke*, dann erzeuge ich lediglich ein mentales Modell des eben beschriebenen Aktes der Bezugnahme. Dieses Ereignis erzeugt eine ganz andere logische und psychologische Struktur.)

Es gibt für jedes informationsverarbeitende System, das ein Selbstmodell erzeugt, immer nur *ein* System, das sich mit Hilfe einer öffentlichen Sprache auf dieses Selbstmodell *als internes* beziehen kann. Sie können sich jederzeit auf das Selbstmodell von TM beziehen und aus der Perspektive der dritten Person eine beliebige Anzahl von Aussagen darüber machen, zum Beispiel: *„Er scheint mir tatsächlich etwas verwirrt zu sein!“* Da es jedoch aus ihrer Perspektive immer schon eine externe Datenstruktur in einem physischen System bleibt, das nicht mit Ihnen identisch ist, können Sie sich niemals auf mein Selbstmodell und seine Zustände *unter der Hinsicht der Internalität* beziehen. Wenn Sie oder ich monologische Sätze des Typs *„Ich bin gerade etwas verwirrt“* äußern, so tun wir dies immer in einer privilegierten Situation: Der Produzent oder Erzeuger beider Repräsentate – des externen wie des internen, des propositionalen wie des analogen –, von denen das eine zur Referenz des anderen wird, sind *identisch*. Mit anderen Worten: Nur *ich* kann die semantische Verknüpfung von psychologischer Selbstbeschreibung und mentaler Selbstmodellierung *unter der Bedingung der Identität beider Repräsentationssysteme* leisten. Nur *ich* kann mich unter dieser Bedingung als *Sprecher* auf einen *Erlebenden* beziehen. Daß dies so ist, wird nur wenigen von uns auf der Ebene theoretischer Reflexion deutlich. Ein mentales Modell dieser eigentümlichen Situation erzeugen wir jedoch alle, als psychologische Subjekte *erleben* wir sie – nämlich immer dann, wenn wir sie herbeiführen. Ich behaupte, daß es genau diese Asymmetrie zwischen innerem Erleben und theoretischem Verständnis ist, die dem Problem der Subjektivität mentaler Zustände zugrundeliegt. Im fünften Kapitel werde ich deshalb auf diesen Punkt noch einmal zurückkommen.

Diese Arbeit ist ein Versuch, die Grundlinien eines neuen, naturalistischen Begriffs des psychologischen Subjekts anzubieten. Mein Ziel war deshalb, plausibel zu machen, wie das unter der Überschrift „Subjektivität“ zusammengefaßte Bündel problematischer psychologischer Eigenschaften im Rahmen einer naturalistischen Theorie mentaler Repräsentation einem tieferen Verständnis zugeführt werden könnte, nämlich durch eine Analyse verschiedener Formen mentaler Modellierung, hauptsächlich Selbstmodellierung. Bevor ich die eigentliche These der vorliegenden Arbeit formuliere, möchte ich einen letzten Versuch unternehmen, den Begriff des „Selbstmodells“ in den Augen meiner Leser mit etwas mehr Gehalt zu füllen. Im nächsten Abschnitt wird es darum gehen, anzudeuten, daß dieser Terminus kein empirisch unplausibler Philosophentraum ist, sondern daß er sich tatsächlich zu einer vielfältigen Verankerung in *empirischen* Theorien des Selbstbewußtseins eignet.

3.2.2 Deviante mentale Modelle des Selbst: Ich-Störungen, halluzinierte Selbste, multiple Selbste und luzide Träume

Ich habe schon in Abschnitt 2.3.2 einige Grenzfälle mentaler Realitätsmodellierung untersucht, um zu einem besseren Verständnis der Behauptung beizutragen, phänomenales Bewußtsein sei aufs engste mit repräsentationalen Gesamtzuständen korreliert. Durch eine solche Analyse von psychischen Nicht-Standardsituationen wollte ich neue Perspektiven auf die Frage gewinnen, was es heißt, daß unser normales Wachbewußtsein und seine Inhalte das Resultat vielfältiger Repräsentationsprozesse in unseren Gehirnen sind. In Analogie zu diesem Unternehmen werde ich nun meine Leser erneut zu einem kleinen Seitensprung in die empirische Psychologie veränderter Bewußtseinszustände einladen – denn grau ist alle Theorie und vielerlei Früchte wachsen auf dem goldenen Baum unseres Innenlebens. Manche von ihnen sind süß, andere bitter – doch lehrreich ist ihr Genuß allemal. Lassen Sie uns deshalb gemeinsam einige Fälle betrachten, in denen die Aktivierung ungewöhnlicher Selbstmodelle zu ungewöhnlichen Formen von subjektivem Bewußtsein und Selbsterleben führt.

1) Zerfallende Selbstmodelle: Ich-Störungen

Beginnen wir mit dem phylogenetisch ältesten Teil des menschlichen Selbstmodells, dem mentalen Körperbild. Wir kennen vielfältige Störungen des phänomenalen Körperschemas. Solche Störungen können wir zum Beispiel beim Einschlafen erleben oder in der Aufwachphase nach einer Vollnarkose. Es gibt außerdem neurologische Erkrankungen (zu denen auch einige der von mir am Beginn des letzten Kapitels angesprochenen Neglekte und Agnosien gehören), in deren Folge Patienten Teile ihres Körpers nicht mehr als Teile *ihres* Körpers erleben – etwa, wenn sie ihr eigenes Gesicht nicht mehr im Spiegel wiedererkennen können, eine Hälfte ihres Körpers nicht mehr waschen und ankleiden oder eine Gesichtshälfte nicht mehr rasieren.

Das körperliche Selbstmodell wird aus Information konstruiert, die überwiegend aus drei Quellen stammt: Dem Gleichgewichtssinn, der visuellen Modalität und der propriozeptiven Eigenwahrnehmung. Ein scharf umgrenzter Ausfall der Propriozeption führt zum Ausfall des *Körpergefühls*, desjenigen Teils unseres phänomenalen Selbstmodells, der uns vielleicht als der sicherste und gewisseste überhaupt erscheint. Oliver Sacks beschreibt den seltenen Fall einer sehr selektiven sensorischen Polyneuropathie, der ausschließlich propriozeptive Nervenfasern betraf.[41]

> Aber am Tag der Operation hatte sich Christinas Zustand weiter verschlechtert. Sie konnte nur stehen, wenn sie dabei auf ihre Füße sah. Ihre Hände „machten sich selbständig", und sie konnte nur etwas festhalten, wenn sie sie im Auge behielt. Wenn sie etwas in die Hand nehmen oder etwas in den Mund stecken

[41] Vgl. Sacks 1987: 69ff. Siehe auch Metzinger 1997.

wollte, griff sie daneben oder schoß über ihr Ziel hinaus, als sei sie nicht mehr in der Lage, ihre Bewegungen zu steuern und zu koordinieren.

Zudem konnte sie kaum aufrecht sitzen – ihr Körper „gab nach". Ihr Gesicht war seltsam ausdruckslos und schlaff, ihr Unterkiefer hing herab, und sogar ihre Stimmlage hatte sich verändert.

„Es ist irgend etwas Furchtbares passiert", stieß sie mit einer geisterhaft dünnen Stimme hervor. „Ich spüre meinen Körper nicht. Ich fühle mich wie verhext – als wäre ich körperlos."[42]

. . .

Christina hörte genau zu, mit einer Aufmerksamkeit, die nur die Verzweiflung hervorbringt.

„Ich muß also", sagte sie langsam, „mein Sehvermögen, meine Augen in all den Situationen einsetzen, in denen ich mich bis jetzt auf meine – wie haben Sie das genannt? – Eigenwahrnehmung verlassen konnte. Ich habe schon bemerkt", fügte sie nachdenklich hinzu, „daß ich meine Arme ‹verliere›. Ich meine, sie seien *hier*, aber in Wirklichkeit sind sie *dort*. Diese ‹Eigenwahrnehmung› ist also wie das Auge des Körpers – das, womit der Körper sich selbst wahrnimmt –, und wenn sie, wie bei mir, weg ist, dann ist es, als sei der Körper blind. Mein Körper kann sich selbst nicht ‹sehen›, weil er seine Augen verloren hat, stimmt's? Also muß *ich* ihn jetzt sehen und diese Augen ersetzen. Hab ich das richtig verstanden?"[43]

Sacks berichtet über den weiteren Verlauf der Erkrankung:

Unmittelbar nach dem Zusammenbruch ihrer Eigenwahrnehmung und noch etwa einen Monat später war Christina so schlaff und hilflos wie eine Puppe. Sie konnte sich nicht einmal selbst aufsetzen. Aber schon drei Monate später stellte ich zu meiner Überraschung fest, daß sie sehr gut sitzen konnte – zu gut vielleicht, zu graziös, wie eine Tänzerin, die mitten in einer Bewegung innegehalten hat. Und bald merkte ich, daß dies tatsächlich eine Pose war, die sie, sei es bewußt oder automatisch, einnahm und aufrecht erhielt, eine gezwungene oder schauspielerhafte Positur, die das Fehlen einer echten, natürlichen Haltung ausgleichen sollte. Da die Natur versagt hatte, behalf sie sich mit einem „Kunstgriff", aber das Gekünstelte ihrer Haltung orientierte sich an der Natur und wurde ihr bald zur „zweiten Natur".[44]

Mit jeder Woche wurde das normale, unbewußte Feedback der Eigenwahrnehmung immer mehr von einer ebenso unbewußten Rückmeldung durch visuelle Wahrnehmung, einen visuellen Automatismus und zunehmend integriertere und flüssigere Reflexabläufe ersetzt. Fand bei ihr eine grundlegende Entwicklung statt? Erhielt möglicherweise das visuelle Modell des Körpers, das Körperbild des Gehirns – das gewöhnlich recht schwach (und bei von Geburt an Blinden überhaupt nicht) ausgeprägt und normalerweise dem propriozeptiven Körperschema untergeordnet ist – jetzt, da dieses propriozeptive Körperschema verlorengegangen war, infolge von Kompensation und Substitution in zunehmendem, ungewöhnlichem Maße Gewicht? . . .[45]

[42] Vgl. Sacks 1987: 71f.
[43] Vgl. Sacks 1987: 74f.
[44] Vgl. Sacks 1987: 77.
[45] Vgl. Sacks 1987: 76.

Die Grenzen des Selbstmodells sind die Grenzen des phänomenalen Selbst.[46] An dem Beispiel der „körperlosen Frau" wird zudem deutlich, was es bedeutet, daß das Selbstmodell ein *multimodales* Modell ist: Es kann um einzelne Modalitäten depriviert werden, aber in manchen Situationen den Verlust von Informationsquellen über eine Verstärkung anderer Kanäle funktional kompensieren. Fällt durch einen Defekt auf der „Hardware-Ebene" ein bestimmtes sensorisches Modul aus, bleibt der entsprechende *phänomenale* Verlust jedoch für die Dauer der Störung bestehen. Das Selbstmodell wird um einen bestimmten qualitativen Aspekt ärmer, obwohl die Steuerfunktion in manchen Situationen dadurch rehabilitiert werden kann, daß auf den Informationsfluß aus den verbliebenen Sinnesmodulen verstärkt zugegriffen wird. Die Psychologie des Systems kann sich dabei jedoch tiefgreifend verändern. In dem tragischen Fall, den ich hier als Beispiel anführe, wurde das phänomenale Loch im subjektiven Erlebnisraum zu einem bleibenden Verlust.

> . . . Infolge des noch immer bestehenden Verlustes der Eigenwahrnehmung hat sie das Gefühl, ihr Körper sei tot, nicht wirklich, gehöre nicht zu ihr – sie ist unfähig, eine Verbindung zwischen ihm und sich selbst herzustellen. Es fehlen ihr die Worte, um diesen Zustand zu beschreiben. Sie muß auf Analogien zurückgreifen, die sich auf andere Sinnesorgane beziehen: „Es ist, als sei mein Körper sich selbst gegenüber blind und taub . . . Er hat kein Gefühl für sich selbst." . . .
>
> . . . „Es ist, als hätte man mir etwas entfernt, etwas aus meinem Zentrum. Das macht man doch mit Fröschen, stimmt's? Man entfernt ihnen das Rückenmark, man höhlt sie aus . . .Genau das ist es: Ich bin *ausgehöhlt*, wie ein Frosch . . . Kommen Sie, meine Herrschaften, treten Sie ein, sehen Sie Chris, das erste ausgehöhlte menschliche Wesen. Sie hat keine Eigenwahrnehmung, kein Gefühl für sich selbst – Chris, die ausgehöhlte Frau, die Frau ohne Körper!" Sie bricht in ein haltloses, fast hysterisches Lachen aus. Ich beruhige sie, während ich denke: Hat sie vielleicht recht?[47]

Der räumlich codierte Teil unseres Selbstmodells, unseres Erlebniszentrums kann *ausgehöhlt* werden. Dann sind wir uns nur noch visuell als *res extensa* gegeben. Was aber geschieht, wenn *höhere* Formen mentaler Repräsentation ausfallen, Formen die intellektuell-kognitive Funktionen in der Ökonomie unseres Innenlebens erfüllen?

---

[46] Das Selbstmodell ist also genau diejenige Partition des von einem Repräsentationssystem intern geöffneten Darstellungsraums, der zu seinem phänomenalen *Innenraum* wird. Thomas Nagel hat in einer frühen Arbeit darauf hingewiesen, daß eine der zentralen philosophischen Fragen auf die inneren Grenzen des Selbst, auf den Horizont des inneren Raumes verweist. Kann uns ein auf der personalen Ebene angesiedelter Begriff des Wissens oder der absichtsvollen Handlung bei dieser kategorialen Demarkation behilflich sein? *„It may be (though I doubt it) that the idea of a person, with which these other concepts are bound up, is a dying notion, not likely to survive the advances of scientific psychology and neurophysiology. Perhaps we shall have to fall back on the idea of an organism or an organic system."* (Nagel 1969: 457)

[47] Vgl. Sacks 1987: 79f.

Amnesien, die als Folge von Verletzungen der Schläfenlappen, chronischem Alkoholmißbrauch, Hirnhautentzündungen, Elektroschocktherapie oder vorübergehendem Sauerstoffmangel auftreten können, berauben die Person um einen mehr oder minder ausgedehnten Teil ihrer Biographie. Ohne auf die verschiedenen Ätiologien und die vielen unterschiedlichen Formen von retrograder Amnesie einzugehen, kann man sagen: Hier handelt es sich um Verluste der *subjektiven Geschichte* eines Menschen. Das System kann gespeicherte Selbstmodelle nicht mehr aktivieren und in sein aktuelles Selbstmodell einbinden, deshalb kann im Extremfall das im Jetzt gefangene psychologische Subjekt seine Historizität nicht mehr erleben. Die Person ist sich darum innerpsychisch nicht mehr als eine *gewordene* gegeben.[48] Fallen zusätzlich Teile von Lernfunktionen mit aus, dann fehlt dem System auch die Möglichkeit, nach dem Trauma gewisse Teile des Selbstmodells aus neuen Erlebnissen neu zu konstruieren und zu speichern, das heißt: in Form physischer Operationen und Strukturveränderungen bleibend zu fixieren. In solchen Fällen schweren Gedächtnisverlustes geht dem Organismus ein Großteil seiner Möglichkeiten verloren, sich selbst mental als eine zeitliche Entität zu repräsentieren. Er kann sich mental nicht mehr als eine innere und eine äußere Geschichte durchlaufend modellieren.

Denken besteht – wie wir bereits gesehen haben – im wesentlichen in der Fähigkeit, mentale Modelle propositionaler Repräsentate in der richtigen zeitlichen Reihenfolge zu aktivieren (das heißt: logische Relationen intern auf zeitliche abzubilden) und Begriffe zu bilden (aus einer gegebenen Menge von präsentierten und gespeicherten Reizsituationen *mentale Prototypen* zu bilden[49]).Im Normalfall werden auch solche Modelle immer schon in das Selbstmodell eingebunden, und daraus resultiert die subjektive Erfahrung des *„Ich denke"*. Wann immer diese subjektive Erfahrung auftritt – dies ist die heutige Formulierung der cartesianischen Einsicht – gibt es ein Selbstmodell und folglich auch irgendein physisches System, das dieses Selbstmodell in sich erzeugt. Falsch ist eine Theorie der Subjektivität, die die innere Erfahrung expliziert, das phänomenale Modell des Selbst sei der Originator und Erzeuger der Gedanken, der internen Modelle logischer Operationen.[50] Das diese Zustände produzierende System ist das Gehirn und nicht das Selbstmodell.

---

[48] Die diachronische Identität der phänomenalen Person hat damit keinerlei Basis mehr, weil die subjektive Kontinuität des Selbsterlebens durch mentale Selbstmodellierung erzeugt wird.

[49] Die Stärke des PDP-Ansatzes für die Erklärung höherer kognitiver Leistungen durch Analogrepräsentation besteht genau darin, daß in seinem Rahmen mit höchster formaler Präzision gezeigt werden kann, wie die „innere Begriffsbildung" durch Aktivierung von Prototypen und auf der Grundlage subsymbolischer Informationsverarbeitung geleistet werden könnte. Vgl. z. B. Churchland 1989, Bechtel/Abrahamsen 1990.

[50] Introspektion ist, wie schon mehrfach betont, eine zweifelhafte Methode des Erkenntnisgewinns. Trotzdem behaupte ich, daß die klassisch-cartesianische Interpretation unserer intellektuellen Aktivität stark theorieinfiziert und keineswegs auch nur intuitiv zwingend ist: *Sorgfältige* Introspektion erzeugt bekanntlich eben gerade *kein* phänomenales Modell des Selbst als einem Initiator durchgängig absichtsgeleiteter Gedankenketten.

Bei *Denkstörungen*, die in Psychosen der verschiedensten Form auftreten, wird das phänomenale Ich intellektuell inkonsistent. Das geschieht dadurch, daß eine gewisse Klasse von mentalen Modellen mehr oder weniger chaotisch und unkoordiniert aktiviert und in das Selbstmodell eingebunden wird. Lesen wir kurz das Protokoll einer experimentell ausgelösten Psychose:

> „. . . da ist ein Pfarrer neben mir gestanden, der halb im Ruhestand ist. Sie haben mich unterbrochen. Die Reden von Pulver sind vorbereitet, der Pfarrer tut Medikamente und Musik vorbereiten. Nehmen Sie noch jemanden dran heute, aha, jetzt dachte ich soeben, es sei Vormittag. Ich bin in der Kappelenbrücke gewesen, habe einen großen Hecht gefangen. (Draußen hört man etwas.) Aha, das ist der Opthalmologe. Er ist gekommen, wegen der Waschmaschine, das ist sicher. Fräulein X. ist im Bad am hellen Tag, sie wollte mich aber nicht beleidigen. Die Mutter hat gesagt, der Sohn sei schlechter geworden in M., und der Pulver wollte den P. nicht beleidigen. Aber Klages sagt immer so pseudowissenschaftliche Sachen, und deswegen ist er zur Graphologie gegangen."[51]

Lockerungen des Assoziationsflusses[52] und formale Denkstörungen, die sich hier so deutlich in den Äußerungen der Versuchsperson widerspiegeln, bestehen in einer internen Überschwemmung des Repäsentationsraumes mit nur noch schwach verknüpften mentalen Modellen. Werden sie, wie häufig in Fällen von Schizophrenie, *nicht* in das Selbstmodell eingebunden, dann entsteht ein erlebtes Realitätsmodell des Typs *„Die Welt ist voller fremder Gedanken"* oder *„Mir werden fremde Gedanken gesendet"*. Findet eine Einbettung der zugrundeliegenden Datenstrukturen statt, dann zerfällt der entsprechende Teil des *intellektuellen* Selbstmodells. Je nachdem wie stark die Störung ist, kann sie von einer noch subjektiv erlebbaren (mental modellierten) und sprachlich kommunizierbaren geistigen Verwirrtheit bis hin zur völligen Desorientiertheit reichen. Ist eine solche Desorientiertheit nicht nur funktional sondern auch phänomenal maximal ausgeprägt, dann kann man sagen, daß solche Personen keine *geistigen Subjekte* im Sinne von selbstbewußten Denkern mehr sind.

Die chaotisch aktivierten mentalen Modelle sind in diesem Fall afunktionale mentale Simulate, weil sie keine Funktion *für* das System mehr erfüllen. Sie stehen in keiner Repräsentationsbeziehung zu internen oder externen Repräsentanda, insofern gilt hier vieles, was ich in Abschnitt 2.3.2 über

---

[51] Es handelt sich hier um ein Skopolamin-Experiment, zitiert nach Heimann 1989 (in Pöppel 1989: 34f.).

[52] Die Lockerung des Assoziationsflusses kann man vielleicht als allgemeine Geschwindigkeitserhöhung innerhalb eines „aufgeheizten" Systems auf dem Weg durch eine Sequenz weniger stabiler Zustände beschreiben. Zwischen einem Hopfield-Netz (vgl. Hopfield 1982) und einem physikalischen System kann man eine Analogie herstellen, die auf einer mathematischen Äquivalenz zwischen dem vom Netzwerk angestrebten Gleichgewichtszustand und dem energieärmsten Zustand in einem thermodynamischen System beruht. Jeder neue Aktivierungszustand eines solchen konnektionistischen Systems stellt einen „energieärmeren" Gesamtzustand des Systems dar; eine Erhöhung des energetischen Niveaus führt zu instabileren Zuständen.

Halluzinationen gesagt habe. Man kann vielleicht sagen: Denkstörungen sind *semantische Halluzinationen*. Natürlich sind die tatsächlichen Fälle pathologischer Selbstsimulation wesentlich komplizierter, zumal die Störungsbilder selten so scharf umrissen sind, wie die von mir ausgewählten Beispiele suggerieren könnten. Zum Beispiel können scheinbar sinnlose Äußerungen von Psychotikern auch mit Konfabulationen durchsetzt sein, das heißt mit den externen Resultaten von verzweifelten Versuchen des Gehirns, alle durch es erzeugten mentalen Modelle (auch die „nicht-intendierten") doch noch zu einem möglichst konsistenten Modell der Welt und des Selbst in ihr zusammenzufügen. Die Konfabulationen zugrundeliegenden mentalen Modelle sind schlechte Modelle, mit denen sich das System seinen schon als pathologisch erkannten Gesamtzustand erklären will.

Betrachten wir nun Fälle, in denen sich das Selbstmodell von Menschen nicht nur teilweise, sondern sehr weitgehend auflöst. Adolf Dittrich hat in einer sorgfältig durchgeführten empirischen Untersuchung[53] an insgesamt 393 Versuchspersonen gute Belege für die Hypothese gefunden, daß es *ätiologie-unabhängige* Strukturen veränderter Wachbewußtseinszustände gibt, die bei durch unterschiedliche Stimuli ausgelösten veränderten Bewußtseinszuständen erhalten bleiben. Zur Auslösung der veränderten Zustände dienten Halluzinogene I. Ordnung, Lachgas, sensorische Deprivation, die Einleitung hypnagoger Zustände, Hypnose, Autogenes Training und Reizüberflutung durch künstlich erhöhte Rhythmizität des Wahrnehmungsfeldes bzw. erhöhte Stimulusvariabilität.

Ein speziell entwickelter Fragebogen für die Auswertung der bei den Experimenten auftretenden und meist nur wenige Stunden andauernden Episoden abweichender repräsentationaler Gesamtzustände wurde eingesetzt, um deren phänomenologische Merkmale zu erfassen. Entsprechend der zentralen Ausgangshypothese des Projekts blieben die Interkorrelationen zwischen den durch einzelne Items erfaßten Merkmalen der durch verschiedene Auslösemechanismen entstandenen veränderten Bewußtseinszustände stabil. Dimensionsanalysen ergaben, daß es eine sekundäre und drei primäre Dimensionen gab, die ätiologie-unabhängig in allen untersuchten Zuständen aufzuweisen waren. Nach Verfahren der klassischen Testtheorie konnten drei zuverlässige, positiv miteinander korrelierende primäre Skalen konstruiert werden, denen – gegenüber den jeweiligen kausalen Antezedentien invariante – phänomenale Grundtypen solcher veränderter mentaler Realitäts- und Selbstmodelle entsprechen.

Vielsagend sind nun in unserem Zusammenhang die nach Kriterien des phänomenalen Gehalts gewählten Benennungen der drei Skalen. Sie lauten *„Ozeanische Selbstentgrenzung"* (OSE), *„Angstvolle Ichauflösung"* (AIA) und *„Visionäre Umstrukturierung"* (VUS). Die Items der zuletzt genannten Skala erfassen solche Veränderungen, die sich im mentalen Modell der ex-

[53] Vgl. Dittrich 1985.

ternen Wirklichkeit ereignen: Optisch-halluzinatorische Erscheinungen[54], auditiv-visuelle Synästhesien (zum Beispiel Item 138: „Die Farben von dem, was ich sah, wurden durch Töne oder Geräusche verändert“)[55], intensives Bedeutungserleben in bezug auf Gegenstände in der Umgebung (man könnte sagen: Die Entstehung eines magischen mentalen Modells der Welt) und eine aus der Erhöhung des allgemeinen neuronalen Erregungsniveaus resultierende „Aktivierung der Vorstellungstätigkeit, daß sie Wahrnehmungscharakter erhält“.[56]

Mit Blick auf die philosophisch interessante Frage nach den Bedingungen des Zerfalls von Subjektivität sind natürlich die Dimensionen *Angstvolle Ichauflösung* und *Ozeanische Selbstentgrenzung* interessant, weil sie unter anderem zeigen, wie die Desintegration des eigenen Selbstmodells von unserem Gehirn emotional sehr unterschiedlich modelliert werden kann. Den klassischen „Horrortrip“, die dysphorische Variante des zerfallenden Subjekts, beschreibt Dittrich wie folgt:

> Darüber hinaus ist die Einheit der Person im Vergleich zum NWB [= normalen Wachbewußtsein, Anmerkung TM] zersplittert, indem die Verfügbarkeit über normale Ichfunktionen nicht mehr gegeben ist: Das Denken ist verändert, indem man Wichtiges von Unwichtigem nicht mehr unterscheiden kann (Item 9), Gedankengänge von Nebengedanken unterbrochen werden (Item 32) und immer wieder abreißen (Item 136). Der Zustand einer „völligen Leere im Kopf“ (Item 158) tritt auf. Das Gefühl der willentlichen Selbstbestimmbarkeit ist verlorengegangen (Item 66), man hat Mühe, auch nur die kleinste Entscheidung zu treffen (Item 107), fühlt sich wie gelähmt (Item 110), wie eine Marionette (Item 83) oder wie ein Automat (Item 55) und verharrt wie erstarrt in einer ganz unnatürlichen Haltung (Item 105).[57]

Interessant in Zusammenhang mit Thomas Nagels *objective self* ist auch Item 157: „Ich beobachtete mich selbst wie einen fremden Menschen“. Hier handelt es sich aber nicht um eine mentale Simulation (ein philosophisches Gedankenexperiment), sondern um einen komplett veränderten *Bewußtseinszustand*, also um einen abweichenden phänomenalen Gesamtzu-

---

[54] Solche mentalen Modelle weisen Formkonstanten über verschiedene repräsentationale Gesamtzustände hinweg auf, die man deshalb vielleicht als determiniert durch Eigenheiten der spezifischen neurobiologischen Implementationsstruktur des menschlichen Gehirns (zum Beispiel der anatomischen Struktur des visuellen Cortex) interpretieren kann. Vgl. Dittrich 1985: 208; Siegel/Jarvik 1975.

[55] In Rahmen der von mir angedeuteten Theorie mentaler Repräsentation bedeutet das, daß mentale Präsentate durch veränderte physische Randbedingungen miteinander interagieren bzw. daß ein sensorischer *information overflow* dazu führt, daß die normale Zuordnung von Format und auslösendem Modul nicht mehr gegeben ist. Vgl. Abschnitt 2.1.3 des vorangegangenen Kapitels.

[56] Vgl. Dittrich 1985: 209. Das zuletzt genannte Phänomen kann man vielleicht durch folgende provisorische Hypothese einer Erklärung näherbringen: Aufgrund des erhöhten Erregungsniveaus gewinnen mentale Simulate – durch die stark angestiegene unterliegende Signalstärke – einen Präsentationsaspekt hinzu und werden deshalb vom System irrtümlich als über die Standard-Kausalketten aktivierte *Repräsentate* interpretiert.

[57] Vgl. Dittrich 1985: 206.

stand, der unter anderem durch eine Störung der Zentrierungsfunktion hervorgerufen wird. Doch davon mehr im nächsten Kapitel; lassen Sie uns zunächst einen Blick auf diejenigen dezentrierten Realitätsmodelle werfen, die gefühlsmäßig vom System allem Anschein nach als in Einklang mit seinen Interessen modelliert werden. Der Zustand der ozeanischen Selbstentgrenzung[58] erinnert mit Blick auf die abendländische Geistesgeschichte an Namen wie Meister Eckehart, Tauler, Seuse oder die heilige Hildegard von Bingen.

> Die Skala „Ozeanische Selbstentgrenzung(OSE)" weist bei extremer Ausprägung auf etwas hin, was als „mystische Erfahrung" bezeichnet werden könnte; die meisten Merkmale der Skala sind wenigstens als Keime solcher Erfahrungen zu verstehen, die einen fließenden Übergang zum normalen Wachbewußtsein aufweisen. Der Begriff „Mystik" wird dabei in einem relativ gut definierten religionswissenschaftlichen Sinne verwendet und nicht als umgangssprachliche Bezeichnung für etwas Diffuses und Unbestimmtes.[59]

Wir stoßen hier auf die Frage nach *nicht-subjektivem* Bewußtsein: Erleuchtung ist keine Erfahrung, weil es in ihr keinen Erfahrenden mehr gibt. Gibt es aber überhaupt stabile, nicht-zentrierte repräsentationale Gesamtzustände ohne wesentliche Einbussen an Funktionalität und kognitivem Gehalt, ohne Regression, Bewußtseinseintrübung oder Deliranz? Könnten wir uns an solche Zustände erinnern, die *niemandes Erlebnisse* waren? Könnten sie *im Nachhinein* – ähnlich wie Träume beim Aufwachen – an das nun aktive, aktuelle Selbstmodell gebunden werden und so zu *meinen* Erlebnissen werden, zu einer aperspektivischen Episode meiner psychischen Biographie? Besitzen Selbstentgrenzungs-Erlebnisse des oben diskutierten

---

[58] In Anlehnung an Rolland, Freud und den dem semi-theoretischen Jargon der Psychoanalyse entlehnten Begriff des „ozeanischen Gefühls" ist diese Skala vielleicht etwas unglücklich benannt. Vgl. Dittrich 1985: 206.

[59] Vgl. Dittrich 1985: 202. Auch ohne das begriffsanalytische Niveau der Religionswissenschaften diskutieren zu wollen muß man daran zweifeln, ob eine echte mystische Erfahrung sich jemals durch einen Fragebogen erfassen lassen wird. Nur drei Items der Skala „OSE" beziehen sich direkt auf das *Selbstmodell* der Probanden: *„Ich fühlte mich, als ob ich schweben würde"* (7), *„Es schien mir, als hätte ich keinen Körper mehr"* (68) und *„Die Grenze zwischen mir selbst und meiner Umgebung schien sich zu verwischen"* (13). Die ersten beiden kann man durch den Ausfall der propriozeptiven Schwereempfindung bzw. als den völligen Zusammenbruch des „körperlichen", räumlich kodierten Selbstmodells interpretieren. Alle *anderen* Items beschreiben eindeutig *subjektive Erlebnisse*, zentrierte phänomenale Zustände (*„Ich fühlte mich ohne äußeren Anlaß sehr glücklich und zufrieden"* ‹84›). Darum bleibt nur Item 13 – die Verwischung der erlebten Ich-Welt-Grenze – übrig, um zu entscheiden, ob solche Zustände (wie die der Skala AIA) auf dem Zerfall eines internen Selbstmodells beruhen, oder ob wir es mit einem hypertrophen, sich bis an die Grenzen der phänomenalen Welt ausdehnenden und sie „kolonisierenden" mentalen Modell des Selbst zu tun haben – wie wir es auch von pathologischen, nicht experimentell ausgelösten Zuständen (zum Beispiel Manien) kennen. Da die Hypothesenarchitektur der zitierten Studie ein anderes Erkenntnisziel verfolgt hat, ist das Merkmal „Subjektzentriertheit/Dezentriertheit" nicht sehr stark in die Formulierung der (die abhängige Variable des Fragebogens „APZ" [Vgl. Dittrich 1985: Anhang I.] bildenden) Fragen eingegangen. Vielleicht können zukünftige empirische Untersuchungen diese Lücke schließen.

Typs einen wie auch immer gearteten epistemischen Gehalt, stehen sie in einer repräsentationalen Ähnlichkeitsbeziehung zur Situation des Systems in der Welt oder sind mystische Erlebnisse zwar die schönsten und ergreifendsten Bewußtseinszustände, die dem Menschen möglich sind, aber aus der Perspektive einer empirisch fundierten, philosophischen Theorie des Geistes leere Makro-Simulate? Solche Fragen beziehen sich auf Bewußtseinszustände, die man im Vorgriff und in Anlehnung auf die in den beiden folgenden Kapiteln diskutierten Gedanken Thomas Nagels als den *Großen Blick von nirgendwo* bezeichnen könnte. Der *Große Blick von nirgendwo* setzt allerdings keinen inneren Perspektivenwechsel voraus, sondern den Tod des phänomenalen Selbst: Das vollständige Zurruhekommen der die Perspektivität überhaupt erst schaffenden Zentrierungsfunktion.

Die klassischen Beispiele für aus zerfallenden Selbstmodellen resultierende Ich-Störungen sind *Schizophrenien*. Es scheint allerdings noch keine wissenschaftlich befriedigende und empirisch abgesicherte Typologie der unter diesem Sammelbegriff zusammengefaßten pathologischen Bewußtseinszustände zu geben.[60] Vielleicht kann eine Neuroinformatik der Zukunft uns präzise Erklärungen dafür anbieten, welche Störungen im Informationsfluß des menschlichen Gehirns verantwortlich sind für die mit der Schizophrenie meist einhergehenden Depersonalisationserscheinungen. Da die Phänomenologie dieser traurigen Zustände weithin bekannt ist, möchte ich sie nicht näher betrachten und wende mich nun einigen Fällen zu, in denen *fiktive* Selbste im inneren Erlebnisraum eines informationsverarbeitenden Systems auftreten.

2) Halluzinierte Selbste

Schon der Bus zum Bahnhof hatte Verspätung und zu allem Ärger haben Sie sich auch noch in der Schlange am falschen Fahrkartenschalter angestellt. Trotzdem erreichen Sie den Zug gerade noch rechtzeitig und lassen sich in einem leeren Abteil erschöpft in den Sitz fallen. Nun betrachten Sie in einem leicht unkonzentrierten und geistesabwesenden Gemütszustand die Passagiere in dem Zug auf der anderen Seite des Bahnsteigs. Plötzlich spüren Sie, wie der Zug, in dem Sie sitzen, sanft, aber mit einer stetigen Beschleunigung anfährt, die Sie in Ihrem Körper spüren. Ebenso plötzlich setzt die körperliche Empfindung aus und Ihnen wird bewußt, daß es der *andere* Zug ist, der langsam begonnen hat, den Bahnhof zu verlassen.

Was Sie soeben erlebt haben war eine kinästhetisch-propriozeptive Halluzination, ein irrtümlich vom Gehirn aktiviertes Modell des Gewichts und der Beschleunigung ihres Körpers. Das in solchen Situationen dominante visuelle Modell ihrer Umgebung ließ in der speziellen Input-Konstellation zwei konsistente Interpretationen zu: Der *andere* Zug fährt an oder der

[60] Auch die Genese dieser endogenen Psychose ist noch weitgehend ungeklärt. Zur Differentialdiagnose und den diagnostischen Kriterien der verschiedenen Formen schizophreniformer Störungen vgl. etwa das DSM-II-R 295.10 – 295.94 (1989: 165ff.).

Zug, in dem *Sie* sitzen, setzt sich in Bewegung. Das visuelle Modell der Realität war unterdeterminiert, die sinnlich gegebene Information unterstützte selbst auf der höchsten Ebene mentaler Modellierung noch zwei gleichwertige Interpretationen. Auf der phänomenalen Ebene sind mentale Modelle in Standardsituationen jedoch immer eindeutig interpretiert (ganz ähnlich wie bei dem bekannten Bild des Neckerschen Würfels, der zu einem gegebenen Zeitpunkt *t* immer nur in *einer* der drei möglichen Interpretationen subjektiv erlebt werden kann). Deshalb hat sich das informationsverarbeitende System, das Sie sind (vielleicht aufgrund einer unbewußten Erwartungshaltung), „entschieden", eines der beiden Modelle zu favorisieren und gleich *ein zu ihm konsistentes, also problemlos in das interne Weltmodell einbettbares, Selbstmodell mitzuaktivieren* – leider war dabei das vom System ausgewählte Weltmodell das falsche. Und deshalb führte die Aktivierung des kinästhetisch-propriozeptiven Selbstmodells zu einer halluzinatorischen Episode. Diese für das Erlebnissubjekt unhintergehbare Episode beruhte darauf, daß durch einen in die Irre gehenden Automatismus kurzzeitig ein mentales Selbstsimulat aktiviert und nicht *als solches* erkannt wurde. Mit der Korrektur des visuellen Modells einige Sekunden nach dem Anfahren des Zuges wird dann aber sofort das zu der phänomenalen Erfahrung eines Körpers, der sanft beschleunigt wird, führende Selbstmodell deaktiviert – und wir stellen amüsiert oder leicht verärgert fest, daß wir uns getäuscht haben.

Dies ist sozusagen der „Minimalfall" einer afunktionalen Selbstsimulation, der hier zu einem partiell leeren oder illusionären Leiberleben führt. Man könnte solche voreiligen Komplettierungen von Realitätsmodellen durch repräsentational passende – das heißt: durch ihre relationale Struktur problemlos in die gegebene höherstufige mentale Struktur einbettbare – Selbstmodelle auch als interne Konfabulationen bezeichnen. Denn es scheint eine der höchsten Prioritäten biologischer Repräsentationssysteme zu sein, auch ambige und unvollständige Datenmengen so schnell wie möglich zu in sich stimmigen mentalen Realitätsmodellen zu komplettieren. Diese biologisch zweifellos nützliche Fähigkeit bringt manchmal Fehlleistungen mit sich, die phänomenal erst im Nachhinein als solche dargestellt werden können.

Es gibt natürlich wesentlich komplexere Selbst-Halluzinationen. Denken wir nur an Träume, in denen wir die bizarrsten psychologischen und physischen Eigenschaften besitzen können, ohne in eine kritische Distanz zu uns selbst zu gelangen.[61] Da das Traum-Ich ein komplexes (wenn auch instabiles) Selbstsimulat ist, das nicht als solches erkannt wird, sondern in einem naiv-realistischen Selbstmißverständnis durch inkonsistente innere Welten taumelt, kann man Träume als Zustände phänomenaler Selbstverlorenheit charakterisieren. Ein skeptischer Phänomenologe könnte dafür argumen-

[61] Im übernächsten Abschnitt werde ich solche Träume diskutieren, die durch das Entstehen einer kritischen Meta-Perspektive von komplexen Halluzinationen zu steuerbaren Pseudo-Halluzinationen geworden sind.

tieren, daß Träume überhaupt nicht-subjektive Bewußtseinszustände sind, weil ein Träumer die Frage *„Wie ist es, ein Träumer zu sein?"* nicht stellen oder beantworten könnte, ohne den Traum zu beenden. Denn die Nagelsche Frage nach der Gesamtqualität des inneren Erlebnisraumes, die durch die sprachliche Formulierung des *„Wie es ist, ein X zu sein"* herausgegriffen werden soll, kann immer erst hinterher, nach dem Wechsel des Realitätsmodells, beantwortet werden. Wie ich von Peter Bieri gelernt habe, bedeutet dies aber nicht, daß Träumer keinen phänomenalen Innenraum besitzen: Es ist letztlich nur eine Aussage über den Gehalt und das funktionale Profil des Selbstmodells im Traum.

Wir kennen sogar noch umfangreichere und auch stabilere Selbst-Halluzinationen des räumlich-körperlichen Selbst als Bewegungsillusionen, Träume oder die von Philosophen so gern diskutierten Phantomglieder.[62] Ein solcher Typus sind die sogenannten *außerkörperlichen Erfahrungen* (AKEs; auch in der deutschsprachigen Literatur häufig als „OBEs" = *Out-of-body-experiences* bezeichnet), bei denen die betreffende Person das unerschütterliche Gefühl hat, ihren physischen Körper (meist in einer Art ätherischen Doubles) zu verlassen und sich außerhalb von ihm zu bewegen. Nicht selten beinhalten diese Bewußtseinszustände ein komplettes visuelles Modell der Welt. Der eigene physische Körper wird dabei von außen – überwiegend aus der Vogelperspektive – wahrgenommen.[63]

Berichte über Erlebnisse dieser Art finden wir in allen Kulturen und in allen Zeiten, deswegen deuten sie auf ein invariantes „neuropsychologisches Potential" des menschlichen Gehirns hin. Man darf annehmen, daß solche Erfahrungen der ebenfalls in vielen Kulturen überlieferten mythischen *Prototheorie des Geistes*[64] zugrunde liegen, nämlich einer Theorie des

---

[62] Neueres empirisches Material deutet darauf hin, daß es eine genetisch determinierte Neuromatrix des Körperschemas gibt, deren kontinuierliche, inputunabhängige Aktivität auch für Schmerzen in Phantomgliedern mitverantwortlich sein könnte. Vgl. Melzack 1989, 1992. – Eine gute Hypothese für die im Text diskutierten OBEs könnte davon ausgehen, daß in speziellen Streßsituationen, in denen das Gehirn von allem propriozeptiven Input abgeschnitten ist, die Aktivität der „fest verdrahteten" Neuromatrix jedoch andauert, ein phänomenales Modell der Form und der Bewegungen des Körpers *ohne Schwereempfindungen* erzeugt wird: Ein ätherischer Leib, mit dem das Erlebnissubjekt durch vom Gehirn simulierte Modelle der Wirklichkeit schweben kann.

[63] Einen kurzen Überblick über die Literatur und die Forschungstrends vom neunzehnten Jahrhundert bis zum Jahre 1987 bietet Alvarado in Alvarado 1989 an. Ein Überblick über die Phänomenologie findet sich in Irwin 1985: 76ff; eine Analyse von verschiedenen Fallstudien in Blackmore 1982a: 56ff, von Berichten über OBEs in anderen Kulturen und verschiedenen wissenschaftlichen Untersuchungen in Blackmore 1982a: 71ff bzw. 82ff.

[64] In vielen Kulturen finden wir vorwissenschaftliche Theorien über einen Lebenshauch (den hebräischen *ruach*, den arabischen *ruh*, den lateinischen *spiritus*, das griechische *Pneuma* oder das indische *prana* bzw. die fünf *koshas* usw.) , welcher den Körper belebt und bei Ohnmachten und im Tod verläßt. Wir haben es hier mit einem noch sinnlich-konkreten Bild des Geistigen zu tun, das *Teile* zusammenhält – auch Gesellschaften und Gruppen von Menschen. In der abendländischen Philosophie des Geistes hat dieser Protobegriff des Geistes durch die Pneumatologie von Anaximenes im sechsten vorchristlichen Jahrhundert über Diogenes von Apollonia und die aristotelische Unterscheidung zwischen Atemluft und psychi-

Geistes als einer noch sinnlich-konkreten Seele in Gestalt einer räumlichen, halbstofflichen Form, die den Körper beim Tod, im Schlaf und bei Ohnmachten verläßt. In den letzten zwei Jahrzehnten sind nun die ersten wissenschaftlichen Theorien über diese Zustände veröffentlicht worden, die testbare Voraussagen machen und sich auf empirische Untersuchungen stützen. All diese Theorien sind letztlich psychologische Theorien, die annehmen, daß es sich bei diesen Erlebnissen um komplexe Halluzinationen handelt. In den drei interessantesten Hypothesen spielt das körperliche Modell des Selbst während einer psychischen Ausnahmesituation die Hauptrolle.

Harvey Irwin geht von einer Verschiebung von Aufmerksamkeitsprozessen bei abgeschwächtem somatosensorischem Input und einer synästhetischen Ergänzung des somästhetischen Bildes durch ein – aus den Datenbanken der Erinnerung aufgebautes – visuelles Modell der Umgebung aus.[65] John Palmer sieht außerkörperliche Erfahrungen als kompensatorische Prozesse nach die Integrität des Selbstmodells bedrohenden Veränderungen des Körperschemas.[66] Susan Blackmore dagegen verwendet sogar explizit den Begriff des „Realitätsmodells", um solche außergewöhnlichen Zustände aus der Perspektive des Informationsverabeitungsansatzes zu erklären. Für sie sind außerkörperliche Erfahrungen episodische Modelle der Realität, die von Gehirnen konstruiert werden, welche in Streßsituationen vom sensorischen Input abgeschnitten sind und auf interne Informationsquellen zurückgreifen müssen (etwa visuelle „kognitive Landkarten" aus dem Gedächtnis, die interessanterweise bei der Mehrzahl der Menschen aus der Vogelperspektive organisiert sind). Dabei entsteht vorübergehend eine sehr realistische – weil einen phänomenalen Körper mit einschließende – visuell komplettierte und erlebnismäßig unhintergehbare mentale Simulation der Welt von einer Perspektive *außerhalb* des physischen Körpers.[67] Es ist noch nicht geklärt, ob es sich bei diesen Zuständen um einen diskreten, durch einen spezifischen Set psychophysiologischer Merkmale charakterisierten Typus repräsentationaler Gesamtzustände handelt oder um eine Sonderform des *luziden Traums* – auf den wir im übernächsten Abschnitt einen kurzen Blick werfen werden. Wenn die eben erwähnten Theorien in die richtige Richtung deuten, dann zeigt dies, wie menschliche Gehirne in ihrem Bestreben, auch unter Bedingungen einge-

---

schem Pneuma (die vielleicht als der erste Naturalisierungsversuch im Okzident gelten darf) eine lange Geschichte durchlaufen. Diese Geschichte führte weiter über alchemistische Theorien der Beherrschung der Natur durch die Beherrschung des *Geistes* und die Neuplatoniker, für die das Pneuma eine die Seele umhüllende Lichtaureole war (die sie vor einer Berührung und Befleckung durch materielle Gegenstände schützen sollte) bis hin zur christlichen Philosophie, die den Geist schließlich personalisierte und denaturalisierte. Die abendländische Geschichte des Geistbegriffs ist somit auch die Geschichte einer bis zum Hegelschen System immer weiter fortschreitenden Differenzierung der traditionalistisch-mythisch-sinnlichen Prototheorie des Geistes.

[65] Vgl. Irwin 1985: 306.

[66] Vgl. Palmer 1978.

[67] Vgl. Blackmore 1984b, 1987.

schränkten Umweltkontakts immer ein möglichst komplettes Realitätsmodell zu konstruieren, deviante phänomenale Modelle des Selbst hervorbringen können. Manchmal beinhaltet dieses neue innere Bild des Ich sogar eine Perspektive auf ein *zweites* Selbstmodell (zum Beispiel den auf einem Operationstisch oder Bett liegenden physischen Körper).

Klassisch-essentialistische Theorien der Subjektivität haben große Schwierigkeit mit der Buntheit und Vielfalt der menschlichen Psychologie, besonders mit Multiplikationen der subjektiven Perspektive innerhalb einer Person. Eine naturalistische Theorie mentaler Repräsentation dagegen erlaubt es prinzipiell, auch auf solche „dissoziativen" Zustände phänomenalen Bewußtseins eine erfolgversprechende Perspektive einzunehmen, ohne unsere Ontologie um Astralkörper oder aggressive Geistwesen bereichern zu müssen. Dieser vorsichtige Optimismus läßt sich auch durch das nun folgende Beispiel abweichender mentaler Selbstmodellierung illustrieren.

3) Multiple Selbste

Ich habe bereits mehrfach darauf hingewiesen, daß mentale Modelle abstrakte Organe sind. Sie besitzen nicht nur phänomenalen und intentionalen Gehalt, sondern auch ein *funktionales Profil.* Dieses funktionale Profil wird durch ihre Rolle in der mentalen Ökonomie des Systems, dessen Organe sie sind, sowie bezüglich einer Umwelt fixiert. Die mentalen Modelle von Biosystemen wurden – genau wie ihre konkreten Organe – im Verlauf der phylogenetischen Evolution funktional optimiert und an sich verändernde Umgebungen (an die jeweilige „kognitive Nische") angepaßt. Funktionale Optimierung bedingt jedoch nicht automatisch eine gleichzeitige epistemische Anreicherung.

Auch das zentrierte Wachbewußtsein menschlicher Wesen ist Resultat der Entwicklung eines besonderen abstrakten Organs – des Selbstmodells – durch biologische Repräsentationssysteme. Vielleicht ist das mentale Selbstmodell des Menschen sogar das komplizierteste und effektivste biologische Instrument, das in der Geschichte unseres Planeten entstanden ist. Selbstverständlich besitzen auch Selbstmodelle ein funktionales Profil *für* die sie benutzenden Organismen *relativ* zu einer gegebenen Umwelt mit ihren spezifischen Problemstellungen. Gute Selbstmodelle sind zum Beispiel unerläßlich für eine erfolgreiche Operation in sozialen Umwelten und für die Interaktion mit anderen Mitgliedern derselben Spezies.[68]

[68] Viele Autoren haben diese Funktion und ihre genetische Bedeutung für das Entstehen subjektiven Bewußtseins betont. Keith Oatley hat diesen wichtigen Aspekt (in Anlehnung an Mead 1964) als *Meadean Consciousness* bezeichnet, als die internalisierte symbolische Repräsentation eines Prozesses, der seine Wurzeln in einem extrapersonalen Beziehungsgeflecht besitzt. Dieser Prozeß definiert das Selbst über soziale Rollen und Regeln, weil er externe Kommunikationssituationen in dialogische Situationen auf der inneren Bühne transformiert. Oatley macht dabei ebenfalls explizit Gebrauch von dem auf Kenneth Craik zurückgehenden

Nun kann ein Mensch im Laufe seines Lebens – vor allem seines *frühen* Lebens – mit „inkonsistenten gesellschaftlichen Datenmengen" konfrontiert werden, also mit zwischenmenschlichen Situationen, die es ihm unmöglich machen, sie durch die Erzeugung eines einzigen inneren Selbstbilds psychisch und funktional zu bewältigen. Die interessante Entdeckung, auf die ich in diesem Abschnitt das Augenmerk meiner Leser lenken möchte, ist nun, daß menschliche Gehirne scheinbar, wenn sie in bestimmten frühen Phasen ihrer Entwicklung mit nicht zu bewältigenden, inkompatiblen sozialen Situationen konfrontiert werden, auch *mehrere* Selbstmodelle „zu verschiedenen Zwecken" erzeugen können. Manchmal erlangen diese multiplen Selbstmodelle eine weitgehende Autonomie und werden zu permanenten Bestandteilen des Innenlebens der betreffenden Person. In unserem theoretischen Zusammenhang sind diese (gar nicht so seltenen und gut dokumentierten) Phänomene deshalb von Interesse, weil sie multiple bzw. *alternierende phänomenale Perspektiven* in ein und demselben Repräsentationssystem mit sich bringen.

Wovon ich hier spreche, sind sogenannte *Dissociative Identity Disorders* (**DID**s). Hierbei handelt es sich um einen Typ von psychiatrischen Störungsbildern, der sehr häufig auf extreme frühkindliche Traumata zurückzuführen ist (in den allermeisten Fällen ist das sexueller Mißbrauch durch ein Elternteil, überwiegend den Vater). Nach offiziellen diagnostischen Definitionen sind Fälle multipler Persönlichkeit dadurch gekennzeichnet, daß innerhalb eines Individuums zwei oder mehr[69] Persönlichkeiten existieren, die zu unterschiedlichen Zeiten dominant werden und dann das Verhalten des Individuums bestimmen. Jede dieser Unterpersönlichkeiten besitzt eine komplexe Struktur sowie eigene, unverwechselbare Verhaltensmuster und soziale Beziehungen. Meistens gibt es eine „Gastgeberpersönlichkeit" (die gewöhnlich amnestisch bezüglich der Episoden ist, in denen andere Persönlichkeiten die Bühne betreten) und eine Reihe von „Alter Egos", die sich normalerweise mit je eigenen Namen benennen. Keine der Persönlichkeiten scheint über ein volles emotionales Spektrum zu verfügen. Häufig ist die Gastgeberpersönlichkeit affektiv undifferenziert, während die Gastpersönlichkeiten ein in verschiedenen Richtungen übertriebenes *affektives Profil* besitzen – das sie dadurch für bestimmte soziale Situationen „geeignet" macht. Die verschiedenen Subpersönlichkeiten scheinen ein allgemeines Hintergrundwissen über die Welt miteinander zu teilen, aber eine eigene Lebenserfahrung und ein spezifisches Selbstgefühl

---

Begriff des „mentalen Modells": *„It seems likely that Vygotskyan and Meadean abilities have evolved from abilites to construct Craikian models (. . .). It may be, however, that although the* ***mechanisms*** *have evolved in this way, what we experience as consciousness, the phenomenology of explicitly knowing, and knowing that we know, derives from the* ***socially*** *derived sense of the self as director and as part of the comparison processes of consciousness. If so, the phenomenology would be affected by, and derivative from, this sense."* (Oatley 1988: 378; vgl. dazu auch Oatley 1985)

[69] Statistisch gesehen, legt sich eine durchschnittliche Anzahl von *elf* Subpersönlichkeiten nahe. Vgl. Dennett/Humphrey 1989: 70.

während der Perioden aufzubauen, in denen sie die Kontrolle über das Verhalten des Individuums erlangt haben.

Daniel Dennett und Nicholas Humphrey (auf die ich mich, was das oben Gesagte angeht und im folgenden, stütze[70]) schildern eine Patientin namens Mary. Sie ist Anfang dreißig, leidet unter Depressionen, Verwirrtheitszuständen und Erinnerungslücken. Sie war bereits mehrfach in Behandlung, spricht auf medikamentöse Therapien nicht an (weshalb sie auch schon für eine Simulantin gehalten wurde) und ist verschiedentlich als schizophren, unter einer *borderline*-Psychose leidend oder als manisch-depressiv diagnostiziert worden. Einem Therapeuten schildert sie ihre Biographie wie folgt:

> Mary's father died when she was two years old, and her mother almost immediately remarried. Her stepfather, she says, was kind to her, although „he sometimes went too far." Through childhood she suffered from sick headaches. She had a poor appetite and she remembers frequently being punished for not finishing her food. Her teenage years were stormy, with dramatic swings in mood. She vaguely recalls being suspended from her high school for a misdemeanor, but her memory for her school years is patchy. In describing them she occasionally resorts – without notice – to the third person („She did this, That happened to her"), or sometimes the first person plural („We [Mary] went to Grandma's"). She is well informed in many areas, is artistically creative, and can play the guitar; but when asked where she learned it, she says she does not know and deflects attention to something else. She agrees that she is „absentminded" – „but aren't we all?": for example, she might find there are clothes in her closet that she can't remember buying, or she might find she has sent her niece two birthday cards. She claims to have strong moral values; but other people, she admits, call her a hypocrite and liar. She keeps a diary – „to keep up," she says, „with where we're at."[71]

Nach einigen Monaten der Behandlung entdeckt der Therapeut, daß sich die Handschrift der Einträge in Marys Tagebuch von Eintrag zu Eintrag so stark unterscheidet wie die unterschiedlicher Personen. Er entschließt sich zu einer Hypnosesitzung mit seiner Patientin, in deren Verlauf er denjenigen Teil von Mary, „der sich noch nicht zu erkennen gegeben hat", auffordert, dies zu tun. Darauf geschieht das folgende:

---

[70] Ein eindrucksvoller autobiographischer Bericht über die Diagnose und Therapie einer DID-Patientin liegt in deutscher Sprache unter dem Titel *„Ich bin Viele"* vor; vgl. Casey 1992. Patientin und Therapeutin schildern in ihm, wie es ihnen im Laufe eines langen therapeutischen Prozesses gelang, einen „Schwarm" von 24 Persönlichkeiten zu integrieren. Im Nachwort weist Frances Howland auf die unterschiedlichen funktionalen Profile alternierender Selbstmodelle hin: *„Untersuchungen der Gehirn- und Nervenfunktionen bei solchen Patienten haben eindeutige wissenschaftliche Beweise für die Realität von MPD [DID; TM] erbracht. Topographische EEGs, IQ-Tests und andere Verfahren zur Messung der kognitiven Performanz, Sehtests sowie Messungen der kardiovaskulären Funktionen ergaben den Untersuchungsberichten zufolge signifikant unterschiedliche Resultate für die alternierenden Persönlichkeiten innerhalb ein und desselben Körpers."* (Casey 1992: 434f) Eine weitere Fallschilderung unter philosophischen Aspekten mit zusätzlichen Literaturverweisen findet sich bei Wilkes 1988: 109ff. Vgl. auch neuere Bemerkungen von Dennett in Dennett 1991: 419ff und Miller & Triggiano 1992.

[71] Vgl. Dennett/Humphrey 1989: 71.

> A sea change occurs in the woman in front of him. Mary, until then a model of decorum, throws him a flirtatious smile. „Hi, Doctor," she says, „I'm Sally. Mary's a wimp. She thinks she knows it all, but I can tell you . . ."
>
> But Sally does not tell him much, at least not yet. In subsequent sessions (conducted now without hypnosis) Sally comes and goes, almost as if she were playing games with Dr. R. She allows him glimpses of what she calls the „happy hours", and hints at having a separate and exotic history unknown to Mary. But then with a toss of the head she slips away – leaving Mary, apparently no party of the foregoing conversation, to explain where *she* has been.[72]

Im Laufe der Behandlung treten nun weitere *Alter Egos* auf: die kokette „Sally", die aggressive „Hatey" und die junge und gefügige „Peggy". Jede dieser Gastpersönlichkeiten besitzt ihre eigene Geschichte und ihre eigenen Erinnerungen. Die Gastpersönlichkeiten beanspruchen zusätzlich ein weitgehendes Wissen bezüglich der Biographie ihrer „Gastgeberin" Mary, während diese bestreitet, mehr als ein indirektes Wissen über *deren* „Erlebnisse" und Persönlichkeitsgeschichte zu besitzen.

Die weitere Therapie ergibt – auf der Suche nach einer Möglichkeit, die verschiedenen phänomenalen Selbste zu fusionieren – das folgende Bild. Im Alter von vier Jahren wurde Mary von ihrem Stiefvater regelmäßig sexuell mißbraucht, wobei er ihr den Kosenamen „Sandra" gab und ihr einschärfte, daß „Daddy-love" das kleine Geheimnis von ihm und Sandra wäre.[73] Als das psychische Leiden und die Situation für das kleine Kind unerträglich wurde und ihre im Entstehen begriffene Persönlichkeit vollends überforderte, versuchte sie sich durch eine *Spaltung des phänomenalen Selbst* zu retten.

> Eventually, when the pain, dirt, and disgrace became too much to bear, Mary simply „left it all behind": while the man abused her, she *dissociated* and took off to another world. She left – and left Sandra in her place.[74]

Durch diese erzwungene episodische Amnesie entstand ein Mary-Selbst mit einer stabilen phänomenalen Identität, einer konsistenten inneren Geschichte und funktionierenden sozialen Beziehungen. Dennett und Humphrey spekulieren, daß das im Hintergrund gehaltene Sandra-Selbst sich nun weiter gespalten haben könnte, indem es verschiedene Aspekte der entsetzlichen Serie von traumatischen Erlebnisse nochmals auf Sub-Selbste verteilte, die allerdings Zugang zu den mit Mary geteilten Erinnerungen hatten. Der Vorteil hätte – aus der „Perspektive" des dissoziierenden Sandra-Selbstes – darin gelegen, daß nun wenigstens *Teile* des in der traumatischen Situation konstruierten Selbstmodells in manchen „sozial adäquaten" Situationen in den Vordergrund treten und die Kontrolle über Marys Verhalten übernehmen konnten.

---

[72] Vgl. Dennett/Humphrey 1989: 72.
[73] Ich erspare meinen Lesern hier die Details; vgl. Dennett/Humphrey 1989: 73.
[74] Vgl. Dennett/Humphrey 1989: 73.

> . . . Thus her experience of liking to please Daddy gave rise to what became the Sally-self. Her experience of the pain and anger gave rise to Hatey. And the experience of playing at being a doll [dieser Ausdruck bezieht sich auf Versuche, sich einfach passiv zu verhalten und sich sozusagen „totzustellen"; Anmerkung TM] gave rise to Peggy.
>
> Now these descendants of the orginal Sandra could, with relative safety, come out in the open. And before long, opportunities arose for them to try their newfound strength in settings other than the original abuse. When Mary lost her temper with her mother, Hatey could chip in to do the screaming. When Mary was kissed by a boy in the playground, Sally could kiss him back. Everyone could do what they were „good at", and Mary's own life was made that much simpler. This pattern of what might be termed „the division of emotional labor" or „self-replacement therapy" proved not only to be viable, but to be rewarding all around.[75]

Natürlich müßte eine wirklich schlüssige phänomenologische Analyse solcher Spaltungszustände wesentlich detaillierter sein und vor allem iatrogene Artefakte ausschließen können. Skepsis gegenüber den Phänomenen und den spezifischen Interessen der Therapeutengemeinschaft sind durchaus angebracht. (Dennett und Humphrey berichten allerdings auch von einer Patientin, deren Skepsis gegenüber der Diagnose ihrer Therapeutin verschwand, als sie feststellen mußte, daß eines ihrer *Alter Egos* sich bereits bei einem anderen Therapeuten in Behandlung begeben hatte.) Andererseits können an der Existenz vieler solcher Fälle multizentrierten Bewußtseins kaum vernünftige Zweifel geltend gemacht werden, wie auch immer die endgültige wissenschaftliche Beschreibung der jeweiligen Ätiologien lauten mag. Eine befriedigende naturalistische Theorie des Geistes muß darum Erklärungen für das Auftreten multipler Erlebnisperspektiven in pathologischen mentalen Modellen der Welt anbieten können. Der in speziellen Streßsituationen für ein natürliches Repräsentationssystem – das gerade erst begonnen hat, ein stabiles, höherstufiges Selbstrepräsentat zu erzeugen – auftretende Zwang zur „emotionalen Arbeitsteilung", auf den Dennett und Humphrey in dem obigen Zitat anspielen, könnte ein Schlüssel zum Verständnis solcher phänomenal-funktionalen Dissoziationen sein. Emotionale Selbstmodelle sind nämlich komplexe Datenstrukturen, die *für* das sie erzeugende System eine möglichst stimmige interne Repräsentation seiner Interessenlage leisten müssen. Wenn in der Phase, in der das phänomenale Selbst eines Kindes sich gerade erst zu konsolidieren beginnt, ein Elternteil unter dem Deckmantel der Zuneigung zum Aggressor wird (der zudem noch eine zweite Identität durch einen Kosenamen *anbietet*), dann entsteht eine bizarre Interessenlage. Da das System nicht in der Lage ist, sich aus der äußeren Situation zu befreien, kann es unter Umständen gezwungen sein, die interne Modellierung seiner Interessenlage auf *mehrere* Selbstrepräsentate zu verteilen, die die nicht miteinander zu vereinbarenden Funktionen für das System als Ganzes *separat* ausüben.

[75] Vgl. Dennett/Humphrey 1989: 74.

Solche Selbstmodelle können sich in einem System stabilisieren und in späteren Stationen seiner Geschichte durch ganz andere, aber strukturell verwandte soziale Situationen wieder aktiviert werden.

Dann entstehen nicht nur Erinnerungslücken und inkompatible Verhaltensmuster, sondern auch psychologische Eigenschaften, die der externe Beobachter – wie im Fall von unter DID leidenden Patienten – nicht mehr konsistenterweise als Eigenschaften *einer* Person beschreiben kann. Dies zeigt uns sehr deutlich, daß die psychologischen Eigenschaften, die eine personale Psychologie dem Gesamtsystem als ihrem logischen Subjekt zuschreibt, in genetischer Perspektive durch komplexe Ereignisse der internen Selbstmodellierung auf der subpersonalen Ebene erklärt werden müssen. Gute Ansätze für solche Erklärungsstrategien gibt es bereits. Egologisch-metaphysischen Theorien des Geistes bleibt dagegen als Ausweg in der Konfrontation mit solchen psychischen Phänomenen nur die Verdrängung des empirischen Materials durch Betrugshypothesen oder durch okkultistische Annahmen, wie etwa die katholische Kirche sie heute noch vertritt: Einer nur mit Hilfe speziell geschulter Exorzisten zu begegnenden Besessenheit durch aggressive Geistwesen.

4) Luzide Träume: *Wach' ich oder träum' ich?*

Abschließend möchte ich untersuchen, was geschieht, wenn Selbstmodelle sich in Bewußtseinszuständen etablieren, die normalerweise nur schwach oder intermittierend zentriert sind, das heißt: Zustände in denen das Selbstmodell normalerweise hochgradig instabil ist. Als Beispiel werde ich einen *simulationalen Gesamtzustand* – eine komplexe mentale Simulation ohne epistemischen Gehalt – wählen, den wir alle kennen und den ich bereits früher auf seine Bedeutung für das philosophische Problem der Subjektivität hin untersucht habe. Ich meine den Traum, der sich in einen *luziden Traum* verwandelt hat. Was genau sind solche luziden Träume ?

Luzide Träume (Klarträume, *lucid dreams*) sind all solche Träume, die die fünf Kriterien des folgenden Minimalkatalogs erfüllen:

(a) Der Träumer ist sich vollständig darüber im klaren, *daß* er träumt. Er *weiß*, daß er sich in einem Klartraum befindet und ist in der Lage, sich diese Eigenschaft selbst zuzuschreiben.

(b) Der Bewußtseinszustand des Träumers ist in keiner Weise eingetrübt. Die allgemeine Bewußtseinsklarheit liegt mindestens auf dem Niveau des normalen Wachzustandes, nicht selten jedoch darüber („*high dreams*"[76]).

(c) Alle fünf Sinne funktionieren dem subjektiven Erleben nach genau so gut wie im Wachzustand.

(d) Es existiert eine uneingeschränkte Erinnerung an das bisherige Wach*und* Klartraumleben. Es gibt keine asymmetrischen Amnesien wie in manchen Fällen multipler Persönlichkeiten.

[76] Vgl. dazu Tart 1972.

(e) Es ist ein klares Bewußtsein der eigenen Entscheidungsfreiheit vorhanden.

Luzide Träume treten sowohl spontan als auch als Folge bewußter Induktion auf; die Fähigkeit zum Klarträumen ist in gewissem Umfang erlernbar. Die Diskussion um den oben angegebenen Kriterienkatalog ist noch nicht abgeschlossen, es scheint sich jedoch abzuzeichnen, daß der luzide Traum einen für die wissenschaftliche Erforschung neuen diskreten Hauptbewußtseinszustand darstellt. Das kognitive und therapeutische Potential von Klarträumen sowie ihr erkenntnistheoretischer Status und insbesondere auch ihre Bedeutung für die Philosophie des Geistes sind dagegen noch weitgehend unklar.

Zur Illustration hier zwei Beispiele ganz gewöhnlicher Klarträume:[77]

> „. . . Ich komme mit dem Bus in W. an. Zwei Bekannte (aus meiner Volksschulzeit), M. und ein Mädchen (N.?) steigen aus. Ich denke, der Bus fahre noch zum Hauptbahnhof, wo ich besser umsteigen kann. Er fährt jedoch in Richtung Z. Als er auch noch an einem Kreisel vorbeifährt, an dem er eventuell noch hätte zurückfahren können, bin ich ärgerlich und wünsche mir, das Ganze solle nicht wahr sein.
>
> Sofort wird mir klar, daß dies ein Traum ist. Da ich weiß, daß es schon spät ist (etwa 9 Uhr), will ich überprüfen, ob ich mich in einem echten Klartraum befinde oder ob ich nur hypnagoge Bilder sehe. Zu diesem Zweck betrachte ich meine Körperstellung. Ich *sitze* im Bus: also bin ich in einem Klartraum. Nun spreche ich eine dickere Frau, die vor mir sitzt, an. Sie ist albern und wirkt etwas ordinär. Ich sage ihr, sie solle mir etwas auf ein Blatt Papier schreiben, und zwar so, daß ich es lesen kann. Ihr ist diese ungenaue Anweisung offensichtlich verständlich. Sie steht schon neben meiner Sitzbank, geht dann aber wieder zurück – als hätte sie mein Vorhaben, sie sollte sich mir gegenüber postieren, verstanden. Sie sagt dabei: ‚Dann merkst Du ja gleich. . .' Weiter sagt sie nichts. Ich spekuliere, was ich gleich merken würde: Daß sie ein eigenes oder daß sie kein eigenes Bewußtsein hat? Sie fragt, ob ich etwas zum Schreiben habe. Ich denke (!) mir dann ein Blatt Papier und ziehe es mit der rechten Hand sozusagen aus meiner linken Tasche oder aus dem Nichts; worauf die Frau bemerkt, ich sei ja ein richtiger Zauberer. Das Blatt ist aber nicht richtig greifbar und verflüchtigt sich wieder. Daraufhin nimmt die Frau ein Handtuch und breitet es über der Rücklehne ihrer Sitzbank aus; offenbar, um darauf zu schreiben. Ich starre gebannt auf das Handtuch (was ich dabei wohl zu sehr fixiert habe) und wache auf."

Ein besonders wichtiger Faktor bei der Entstehung von Klarträumen ist die Aufrechterhaltung eines *kritischen Bewußtseins* gegenüber der Realität. Hier der Bericht einer jungen Frau über ihren ersten Klartraum:

> „. . .Ich treffe K., der schon Klarträumer ist, auf der Toilette im Theater. ‚Der kommt mir gerade recht!' denke ich. Sowieso bin ich sauer auf ihn. Da mache ich seit Wochen schon ganz brav meine Übungen und habe noch keinen einzigen Klartraum gehabt. Nicht mal einen ganz kleinen, ganz kurzen! Ich mache

[77] Sie sind einem Buch von Paul Tholey entnommen; vgl. Tholey 1987: 232f & 46f.

ihm Vorwürfe: ‚Du hast mir alles falsch erklärt!' Ich bin völlig fertig mit den Nerven. Und überhaupt, jetzt steigere ich mich in richtige Wut: ‚Das ist doch alles gelogen! Sowas gibt's doch gar nicht! Klarträumen! Ha! Aber mit mir nicht! Jetzt ist Schluß! Ich laß mich doch nicht länger von dir verarschen! Von dir schon gar nicht!'

K. reagiert überhaupt nicht. Er steht vor dem Spiegel und streichelt seine Bartstoppeln. Ich gerate in eine derartige Wut, daß ich ihm von hinten auf den Kopf haue. K. dreht sich lachend um, schaut mir dabei in die Augen. ‚Warum lacht er bloß?' denke ich kurz, dann geht er an mir vorbei in den Spiegel. Ich schreie vor Wut und werfe die Seife nach ihm, treffe aber nicht. Ich war inzwischen so in Rage, daß ich einfach nicht gecheckt habe, daß solche Sachen natürlich nur im Traum vorkommen. ‚Nur Mut, Baby!' sagt er, immer noch lachend. ‚Du kommst schon noch drauf', dreht sich um und verschwindet.

Ich platze fast! Da geht hinter mir die Tür auf, und K. kommt mit einem anderen Mann rein. Eng umschlungen! Beide grinsen mich frech an. Ich bücke mich schnell nach der Seife, da wird mir schwindlig. ‚Das darf doch nicht wahr sein', denke ich, als ich wieder hochkomme und die beiden da so stehen sehe. ‚Der ist doch nicht schwul', der doch nicht! Das wüßt' ich aber! Warum freut er sich nur so?' Plötzlich wird mir eiskalt. ‚Was, wenn das alles ein Traum ist??? Träum' ich oder wach' ich? Was ist denn alles passiert bis jetzt? Das ist doch lächerlich! Ist das ein Traum?' frage ich die beiden. Beide schütteln den Kopf und lachen ganz irre. Egal! K. interessiert mich überhaupt nicht mehr.

Ich denke: ‚Das kann, kann, kann nur ein Traum sein! Das ist ein Traum! Ein Klartraum! Was mach' ich denn jetzt bloß?' K. und der andere sind nicht mehr da. Die Kachelwand faßt sich unheimlich echt an. Kühl und glatt. Ich überlege fieberhaft, was ich jetzt machen könnte. Ich bin wahnsinnig aufgeregt. Ich muß jetzt was machen! Da fällt mir der Film mit Heinz Rühmann ein. Der Mann, der durch die Wand gehen konnte. Das war schon immer mein Traum. Bei ‚Traum' muß ich lachen. Gott, bin ich albern, denke ich. Ich versuche, durch die Kachelwand zu gehen. Die Hand ist drin! Entschlossen gehe ich los. Ich dringe in die Wand ein und muß schon wieder lachen. Es ist herrlich, warm und dunkel. Irgendwie rötlich. Jetzt bin ich durch! Und stehe im Wohnzimmer meiner Eltern! Mutti und Vati trinken Kaffee und öden sich an. Mutti kommt sofort auf mich zu und macht mir Vorwürfe. ‚Was fällt dir ein, einfach so ohne Anmeldung hier durch die Wand zu kommen!' ‚Ach hör auf', sage ich glücklich, ‚du bist doch bloß 'ne Traumfigur . . .'"

Zu den das Auftreten von luziden Träumen begünstigenden Bedingungen zählen ein hohes körperliches Aktivitätsniveau[78] sowie erhöhte affektive Erregung während des Tages.[79] Vor der betreffenden REM-Phase[80] stattfindende Schlafunterbrechungen mit kurzen Aktivitäten im Wachbewußtsein erhöhen ebenfalls die Wahrscheinlichkeit für das Auftreten von Luzidität.[81]

Bei Polysomogrammanalysen von Versuchspersonen, die den Beginn ei-

[78] Vgl. Gackenbach, Curren & Cutler 1983; Garfield 1975.

[79] Vgl. Gackenbach, Curren & Cutler 1983; Sparrow 1976.

[80] REM = *Rapid eye movements*; schnelle Augenbewegungen sind das deutlichste physiologische Korrelat von Traumphasen.

[81] Vgl. Garfield 1975; LaBerge 1980a; LaBerge 1980b; Sparrow 1976.

nes luziden Traums durch vorher verabredete Okularsignale anzeigten[82], ergaben sich als elektrophysiologische Korrelate die ersten zwei Minuten einer REM-Phase, kurze Einbrüche von Wachbewußtsein während einer REM-Phase und/oder erhöhte phasische REM-Aktivität. Die Vermutung liegt nahe, daß eine episodisch überschießende Erhöhung des kortikalen Erregungsniveaus eine der wichtigsten physiologischen Bedingungen für das Entstehen von Luzidität ist. Auf phänomenaler Ebene entsprechen ihr erhöhte Angst oder Streß im Traumgeschehen, die Entdeckung von Widersprüchlichkeiten in der Traumwelt oder die Bewußtwerdung einer „traumartigen" Qualität der geträumten Realität. Es gibt vereinzelte Berichte über das Entstehen von Klarträumen im NREM-Schlaf, jedoch keine gesicherten Erkenntnisse.[83]

Derzeit werden Klarträume experimentell eingesetzt, um Schlafstörungen aufzulösen und Alpträume zu verhindern oder zu beenden. Wenn es dem Träumer gelingt, ein stabiles Bewußtsein bezüglich des eigenen Zustandes zu erlangen und aufrechtzuerhalten, kann er bedrohlichen Gestalten furchtlos entgegentreten. Nun kann er sie vernichten oder auch in Freunde, Helfer oder Informationsquellen transformieren. In anderen Kulturen und auch von westlichen Psychotherapeuten sind diesbezüglich detaillierte Ethiken und Handlungsstrategien entwickelt worden.[84] Gewisse therapeutische Ansätze versuchen über die Kommunikation und Auseinandersetzung mit Traumfiguren Zugang zu unbewußten Persönlichkeitsanteilen zu finden und neue Persönlichkeitselemente zu entwickeln. Als leicht und gefahrlos handhabbarer veränderter Bewußtseinszustand bietet der luzide Traum dem Forscher außerdem die Möglichkeit, die Randbedingungen und Möglichkeiten für den Übergang in *andere* veränderte Bewußtseinszustände (zum Beispiel OBEs[85]) im Selbstversuch zu untersuchen. Der Klartraum wird bereits eingesetzt, um Leistungssportlern das gefahrlose Einüben riskanter Bewegungsabläufe zu ermöglichen und auch, um Astronauten auf den Zustand der Schwerelosigkeit vorzubereiten. Eine interes-

[82] Man macht sich hierbei die Entdeckung zunutze, daß bewußt durchgeführte Augenbewegungen im Traumkörper recht zuverlässig mit analogen Rollbewegungen der Augäpfel des physischen Körpers im Schlaflabor korreliert sind. Es sollte in diesem Zusammenhang nicht übersehen werden, daß es sich hierbei nicht nur um einen ingeniösen Versuchsaufbau handelt, sondern auch um die Erzeugung eines menschheitsgeschichtlich neuen und philosophisch nicht uninteressanten Phänomens: Der interpersonalen Kommunikation *zwischen verschiedenen Hauptbewußtseinszuständen*.

[83] Als einführende Literatur eignen sich LaBerge 1982 und LaBerge 1988 in: Gackenbach/LaBerge 1988.

[84] Vgl. Dentan 1988; Stewart 1972a; Tholey 1987.

[85] OBEs *(Out-of-body-experiences)* oder *„Außerkörperliche Erfahrungen"* sind solche Zustände, bei denen die betreffende Person der unerschütterlichen Überzeugung ist, sich außerhalb ihres physischen Körpers zu befinden, und zwar zumeist in einer Art ätherischen Doubles. Außerdem sind die letzten vier Kriterien des anfänglich genannten Kriterienkatalogs erfüllt. Dies ist ein Grund dafür, daß die Diskussion darüber, ob es sich bei außerkörperlichen Erfahrungen um einen diskreten veränderten Bewußtseinszustand handelt oder um eine Subklasse luzider Träume, mit Sicherheit noch einige Zeit andauert. Vgl. Blackmore 1982a, Palmer 1978, Irwin 1985, Tholey 1987.

sante Frage ist, ob es eine spezifische Anwendung von Klarträumen für den Philosophen oder Wissenschaftstheoretiker gibt. Bis zum heutigen Zeitpunkt war alle Philosophie *Philosophie im Wachzustand*. Könnte es eine *Zustandsspezifität der Philosophie*[86] im Sinne einer besonderen Klasse von Problemen geben, die nur innerhalb eines bestimmten Bewußseinszustandes gesehen und gelöst werden können?

Das Phänomen des Klarträumens ist seit Jahrtausenden bekannt. In den Brennpunkt seriösen wissenschaftlichen Interesses rückte es erst während der letzten zwei Jahrzehnte. Als Anfangspunkt dieser Entwicklung kann vielleicht Celia Greens 1968 erschienenes Buch gelten.[87] Die ersten beiden Dissertationen zum Thema erschienen in den Jahren 1978 und 1980.[88] Man kann sagen, daß sich die Erforschung des luziden Träumens noch in einem präparadigmatischen Stadium befindet – ein Umstand, der diesem psychischen Phänomen die erhöhte Aufmerksamkeit von Wissenschaftstheoretikern und an den metatheoretischen Aspekten der *Cognitive Science* interessierten Philosophen eintragen sollte.

Wenn wir luzide werden und ein gewöhnlicher Traum sich in einen Klartraum verwandelt, entsteht ein psychologisches Subjekt: Es gibt nun einen *Erfahrenden*, der um sich selbst, seinen Bewußtseinszustand und seine Handlungsfreiheit weiß. Dies betrifft besonders die Interaktion mit Traumfiguren: Sind mir im Klartraum begegnende Traumfiguren autonome *Personen*, insbesondere wenn sie dies behaupten und rational dafür argumentieren? Diese Frage allein aus der Perspektive des Wachzustandes zu beantworten, könnte eine schwer zu begründende Form des Chauvinismus sein. Putnam hat bereits 1964 die Frage diskutiert, ob man nicht-biologischen aber psychologisch dem Menschen weitgehend isomorphen Systemen Bürgerrechte zugestehen sollte.[89] Er hatte dabei Roboter im Sinn, also funktionale Systeme, die in Form von Artefakten instantiiert sind. Setze ich einen funktionalen Materialismus als Hintergrundannahme über die Natur mentaler Zustände[90] voraus, so könnten Traumfiguren, die mir im

[86] Vgl. hierzu Tart 1975.

[87] Vgl. Green 1968; einen kurzen Literaturüberblick von Aristoteles bis zu diesem Zeitpunkt gibt LaBerge 1988.

[88] Vgl. Gackenbach 1978 und LaBerge 1980; diese beiden Autoren zählen auch zu den Herausgebern der ersten provisorischen Fachzeitschrift, dem seit 1981 erscheinenden Lucidity Letter.

[89] Vgl. Putnam 1964,1975; siehe auch Metzinger 1985: 192ff.

[90] Vgl. Putnam 1981; zur Problematik der oben angesprochenen Hintergrundannahme mit Blick auf eine Psychologie der propositionalen Einstellungen vgl. Putnam 1991. Dort schreibt Putnam über die Unmöglichkeit einer befriedigenden funktionalistischen Theorie des intentionalen Gehalts mentaler Zustände: *„Doch wenn es darum geht, eine Theorie aufzustellen, die die propositionalen Einstellungen, die semantischen Begriffe mit Bezug auf alle möglichen Spezies erklärt, dann stehen wir dem Problem gegenüber, daß unsere Theorie die möglichen Begriffsbildungsweisen aller physikalisch möglichen Wesen „überblicken" muß. . . . Die Idee der Suche nach einer berechenbaren (oder auch nur wohldefinierten) Äquivalenzrelation zwischen funktionalen Zuständen, die der durch die Praxis der „guten Interpreten" implizit definierten Äquivalenzrelation entspricht, stößt in der bodenlosen Endlosigkeit der Gesamtheit der („im Prinzip") zu interpretierenden Begriffsschemata auf eine unüberwindliche Schwierigkeit."* (Putnam 1991: 166, 167)

Klartraum begegnen und mich vielleicht in erkenntnistheoretische Grundsatzdebatten verwickeln, funktionale Subsysteme – sozusagen diskrete virtuelle Maschinen – sein, die durch Teilmengen *meiner* Hirnzustände realisiert sind.

Angenommen, ich begegne im Klartraum einer *scientific community*, die daran interessiert ist, mich von der fundamentalen Falschheit einiger meiner Theorien über die Traumwirklichkeit zu überzeugen: Darf ich die intersubjektive Verifikation „wissenschaftlicher" Hypothesen *im Traum* und *durch Traumfiguren* akzeptieren? Es könnte durchaus sein, daß innerhalb der Traumwelt der Energieerhaltungssatz der Physik nicht gilt und daß die besten wissenschaftlichen Theorien der Traumwissenschaftler dies bestätigen. Ich wäre somit als experimentierender Epistemologe und philosophischer Psychonaut nicht in der Lage, zu beweisen, daß die Referenten einzelner Terme der Traumphysik nur Elemente eines mentalen Modells sind, welches von einem Gehirn konstruiert wird, das ich üblicherweise als meines zu bezeichnen pflege. Dadurch, daß ich den Prozeß der Erzeugung und Überprüfung von Hypothesen durch mein Aufwachen abbreche bzw. verlasse, habe ich im wahrsten Sinne des Wortes niemandem etwas bewiesen und keinerlei Erkenntnisfortschritt erzielt.

Mit der Luzidität begegnet dem Philosophen ein bisher kaum beachtetes psychisches Phänomen, das man auch als die psychologische Eigenschaft der *Zustandsklarheit* bezeichnen kann. „Luzidität" als Begriff könnte man auf mehreren Ebenen explizieren.

(1) Ein *Träumer* ist genau dann ein luzider Träumer bzw. ein Klarträumer, wenn er sich während des Traumes bewußt ist, *daß* er träumt und wenn er weiß, daß sein gegenwärtiger Bewußtseinszustand die eingangs genannten Kriterien erfüllt.

(2) Ein *Traum* ist genau dann ein luzider Traum bzw. ein Klartraum, wenn in ihm ein stabiles *Zustandsbewußtsein* mitgegeben ist und wenn er von dem Wissen um die Erfüllung der eingangs genannten Kriterien begleitet wird.

(3) Ein *Bewußtseinszustand*[91] ist genau dann luzide, wenn die sich in ihm befindende Person[92] sich der Tatsache, *daß* sie sich in diesem Bewußtseinszustand befindet, bewußt ist und wenn sie weiß, daß sie sich in eben jenem Bewußtseinszustand befindet. Luzide sind repräsentationale Gesamtzustände dann, wenn sie durch ein Selbstmodell zentriert werden, zu dessen

[91] Wenn man, wie ich es hier tue, den aus der phänomenologischen Analyse des Klartraums gewonnenen Begriff der „Luzidität" auf Bewußtseinszustände und mentale Realitätsmodelle im allgemeinen hin generalisiert, dann kann man die der ursprünglichen psychologischen Eigenschaft der Luzidität entsprechenden umfassenderen psychologischen Eigenschaften als Zustandsklarheit und Selbsttransparenz bezeichnen. Bei den diesen generalisierten Eigenschaften entsprechenden Formen des Wissens handelt es sich aber nicht um propositionales, sondern um analoges *Erlebniswissen*.

[92] Schwierig wird es in all jenen Fällen, bei denen eine physische Person mehrere phänomenale Personen benutzt. Können sich auch *Mengen von phänomenalen Personen* in einem Bewußtseinszustand befinden, etwa wenn – wie in Fällen multipler Persönlichkeiten – ein Gehirn gleichzeitig mehrere Selbstmodelle konstruiert?

Gehalt Information über den Typ von Zustand gehört, den sie instantiieren. Luzide ist ein *Realitätsmodell* genau dann, wenn in ihm die Tatsache, daß es ein *Modell* der Wirklichkeit ist, und der *Typ* von Realitätsmodell, zu dem es gehört, mitrepräsentiert ist. In diesem Sinne ist unser normales Wachbewußtsein ein *nicht-luzides* Realitätsmodell.

(4) Eine *Person* ist genau dann luzide, wenn sie sich bewußt ist, *daß* sie sich in einem bestimmten Bewußtseinzustand befindet und wenn sie *weiß*, daß sie sich in eben jenem Bewußtseinzustand befindet.

Da die empirische Erforschung des Phänomens „Klartraum" sich noch in seiner Anfangsphase befindet, ist die philosophische Analyse in Ermangelung einer Theorie[93], deren Metatheorie sie sein könnte, auf phänomenologische Spurensuche angewiesen. Prominente Philosophen der analytischen Tradition haben sich in der Vergangenheit bereits mit der Frage auseinandergesetzt, ob Träume allein aus begrifflichen Gründen als Erlebnisse gelten können oder nicht.[94] Für eine philosophische Oneirologie besteht dagegen heute das Problem, daß sie ein sich ständig vergrößerndes empirisches Wissen über das Phänomen „Traum" nicht mehr ignorieren kann, sondern interpretierend in ihre Theorien integrieren muß.

Nicht-luzide Träume sind keine Erlebnisse in einem starken Sinne, weil sie – während sie sich ereignen – nicht meine Erlebnisse sind. Der normale Träumer gleicht nicht selten einem Betrunkenen, der sich in eine Geisterbahn verirrt hat: Ohne zu wissen, wer er ist, und ohne daß ihn auch die gröbsten Widersprüchlichkeiten und Inkongruenzen innerhalb der Traumrealität dazu veranlassen, dieser gegenüber eine kritische Haltung einzunehmen, taumelt er von einer Verwunderung ins nächste Entsetzen ohne jemals zum *Subjekt* zu werden. Die bizarren Episoden des Nachtlebens verwandeln sich nur dann in Elemente unseres Lebens, wenn wir uns beim Aufwachen (beim Überwechseln in einen anderen repräsentationalen Gesamtzustand) an sie erinnern. So nicht der Vorhang des Vergessens schon beim Aufwachen fällt, versuchen wir – nachdem der Übergang in das viel stärker input-determinierte Modell des Wachzustandes sich ereignet hat – kopfschüttelnd zu verstehen, wie uns so etwas widerfahren konnte. Dabei ist es uns gar nicht widerfahren! Was uns widerfährt, ist – genau betrachtet – das plötzliche Auftreten von *Erinnerungen* an mentale Zustände, in denen es zwar einen phänomenalen Innenraum, aber kein stabiles psychologisches Subjekt gab.

Dies mag wie eine Version der Dennettschen „Cassetten-Theorie" der Traumbewußtheit erscheinen[95], bei der die Prozesse der Komposition und Aufzeichnung völlig unbewußt sind und erst beim Aufwachen in Erlebnisse transformiert werden. Dahinter steht jedoch eine wesentlich schwierigere Frage: Sind nicht-luzide Träume subjektiv im Nagelschen Sinne?[96] Tritt im

[93] Vgl. jedoch Blackmore 1988.
[94] Vgl. Malcolm 1959; und auch die Kritik von Putnam 1964 und Dennett 1981.
[95] Vgl. Dennett 1981.
[96] Vgl. Nagel 1974(1981).

nicht-luziden Traum die Qualität des *„Wie es ist, ein X zu sein"* auf; mit anderen Worten: *Ist es irgendwie*, nicht-luzide zu träumen oder kann man nur sagen, daß es *irgendwie ist, Erinnerungen* an einen nicht-luziden Traum zu haben? Da Träume private Ereignisse sind, sind wir zur Beantwortung auf die Berichte von Träumern angewiesen. Es ist aber logisch unmöglich, daß ein Träumer diese Frage untersucht oder beantwortet, ohne seinen Bewußtseinszustand zu transformieren: Entweder er wacht auf oder er wird zum Klarträumer.

Was aber sind dann normale, nicht-luzide Träume? Ihr epistemischer Gehalt scheint gegen Null zu gehen – zumindest was die aktuelle Umwelt des träumenden Systems angeht. Als ein zweiter Typ von Wirklichkeitsmodell, der durch einen eigenen Set physiologischer Korrelate charakterisiert ist, scheinen sie nur sehr schwach durch externen Input, das heißt: durch den Informationsfluß aus den Sinnesmodulen, in ihrem Gehalt determiniert zu sein. Wenn auch Träume abstrakte Instrumente des Organismus sind, welche Funktion haben sie dann inne? Einige prominente Philosophen[97] haben die Auffassung vertreten, daß Träume weder Funktion noch Bedeutung besitzen und eine Reihe empirischer Forscher halten den Traum entweder für das Resultat von Zufallsprozessen auf der neurophysiologischen Ebene[98] oder auch für eine neurologisch zu beschreibende Fehlanpassung.[99] Andererseits liefert ein so komplexes psychisches Phänomen wie der Traum das perfekte Rohmaterial für nicht überprüfbare Interpretationen jeder Art (seien es nun die Homunkuli der Psychoanalyse oder die Jenseitswelten esoterischer Geheimlehren) und bietet sich deswegen für ideologischen Mißbrauch an. Wenn sich über die Funktion des Traumzustandes für den Organismus derzeit nichts genaues sagen läßt, gilt dies wohl vorläufig auch für den Klartraum.[100] Jedoch kann man im Sinne einer meta-phänomenologischen Analyse fragen: Was genau verändert sich in der Struktur des vom Gehirn errechneten Realitätsmodells, wenn aus einem nicht-luziden Traum ein Klartraum wird?

Beginnen wir mit dem, was sich nicht verändert: Das Modell der externen Realität bleibt im wesentlichen unverändert. Es ist komplex, instabil, nicht selten bizarr und voller Inkonsistenzen.[101]Das *Selbstmodell* dagegen

---

[97] Vgl. Dennett 1981; Fodor 1981a; Malcolm 1959.

[98] Vgl. Hinton/Sejnowski 1986; Hobson/McCarley 1979; McCarley 1983.

[99] Vgl. Crick/Mitchison 1983.

[100] Daß er zumindest eine psychohygienische Funktion haben könnte, legt die Beobachtung der *Senoi*–Traumkultur nahe. In manchen nicht-westlichen Kulturen ist nämlich aufgrund anderer ontologischer Grundannahmen die Fähigkeit zum Klarträumen weit besser entwickelt worden. Das wichtigste Beispiel sind hier die *Senoi* in Malaysien (vgl. Stewart 1972a). Besonders interessant erscheint, daß die in diesen Gesellschaften gut entwickelte Traumhygiene möglicherweise mit dem Phänomen zusammenhängt, daß es in dieser Kultur kaum Geisteskrankheiten, asoziales Verhalten und Kapitalverbrechen gab und gibt. Die Einschätzung der Senoi-Traumkultur ist allerdings stark umstritten; vgl. Dentan 1988.

[101] Eine vorsichtige neurologische Vermutung über das Zustandekommen dieser Situation könnte lauten: Träume beginnen immer dann, wenn das allgemeine Erregungsniveau des

verändert sich sehr stark, in gewissem Sinne entsteht es im Augenblick des Luzidewerdens überhaupt erst. Im normalen Traum wissen wir nicht, wer wir sind, wir kennen unseren Namen nicht, wissen meist auch nicht, an welchem Ort wir uns befinden und welcher Tag heute ist. Häufig handeln wir nicht, sondern reagieren nur auf eine sich ständig verändernde Umwelt voller Überraschungen und Bedrohungen. Wenn wir Glück haben, kommt es dazu, daß die Trauminhalte so bizarr werden, daß wir „gezwungenermaßen" eine Zustandskontrolle durchführen („Das *kann* doch nicht wahr sein . . .!"). Oder ein gut bekannter und sich wiederholender Trauminhalt aktiviert unser Gedächtnis und uns fällt ein, wer wir im *Wachleben* sind („Wieso ist der Universitätspräsident eigentlich mein HiWi?"). Oder das Motiv des Traums selbst taucht durch einen glücklichen Zufall im Traum auf („Jetzt sitze ich schon eine halbe Stunde hier im Kino und warte darauf, daß der Film über die Traumzeit der australischen Ureinwohner anfängt! Irgend etwas stimmt doch hier nicht . . ."). In jedem Fall werden die Datenbanken des Wachlebens wieder zugänglich und damit wird das Selbstmodell stabil. Es wird aber nicht nur durch ein Mehr an Information stabilisiert, es wird auch inhaltlich verändert: Es ist ab jetzt die Repräsentation des Selbst *als* eines Träumers oder einer Träumerin – obwohl auch das nicht notwendigerweise so sein muß.[102]

Die philosophische Interpretation dieses Prozesses wird entscheidend von der zugrunde gelegten Theorie über die kausale Relation zwischen mentalen und neurophysiologischen Ereignissen abhängen. Nimmt man im Sinne eines *Minimal Materialism* an, daß mentale Vorgänge supervenient[103] gegenüber physischen Prozessen sind (daß es also nicht möglich ist, präzise gesetzesartige Korrelationen über die Kluft zwischen der subpersonalen und der personalen Beschreibungsebene hinweg zu formulieren, daß aber dennoch alle psychologischen Eigenschaften eines Systems feststehen, wenn auch seine biologischen Eigenschaften fixiert sind[104]), dann ist es natürlich nicht der *Träumer*, der Zustandsklarheit erzeugt indem er die Datenbanken der Erinnerung öffnet. Da es „Verursachung von oben" im

---

Gehirns bei gleichzeitiger Blockade der Sinnesmodule ansteigt, aber für das Entstehen normaler Erinnerungsleistungen noch nicht ausreicht. Die in einer solchen Situation entstehenden Modelle können sehr komplex (d. h. erlebnismäßig realistisch) sein, sind aber instabil (d. h. die Traumrealität verändert sich häufig, plötzlich und unvorhersehbar) und können nur schwer mit früher aufgetretenen Modellen verknüpft (erinnert, wiedererkannt) werden.

[102] Beim sogenannten „falschen Erwachen" kann auch ein Modell des Selbst als gerade aus dem Traum in den Wachzustand übergegangenem erzeugt werden. Wenn es nun nicht zu einem Rückfall in einen gewöhnlichen Traum kommt, führt die Instabilität der Traumrealität meist schnell ihrerseits zu Luzidität bzw. einem „wirklichen" Erwachen. Wenn man allerdings erst einmal geträumt hat, *daß* man wach ist, dann kann man sich in einem zukünftigen Klartraum an diese Erfahrung erinnern. Das kann zu der unangenehmen Frage führen, ob man jetzt gerade vielleicht nur *träumt, daß man klarträumt*.

[103] Ich habe Vor- und Nachteile der Supervenienz-These an anderer Stelle diskutiert und möchte mich hier nicht auf diese Position festlegen, obwohl sie derzeit als eine der vielversprechendsten Optionen im psychophysischen Problemkreis gelten darf. Vgl. Metzinger 1985, 1990.

[104] Vgl. Kim 1974, 1978, 1982; Haugeland 1982; Stoecker 1992.

Sinne eines personalen Eingriffs in die neuronale Maschinerie diesem theoretischen Modell zufolge nicht gibt, ist es vielmehr das *Gehirn*, welches das luzide Subjekt konstruiert. Aufgrund veränderter Randbedingungen ist das System in der Lage, eine wesentlich reichere Selbstrepräsentation zu erzeugen und kann so eine Kontinuität zu früheren Zuständen erkennen. Damit ist es auch in der Lage, sich selbst *als jetzt gerade in einem bestimmten Zustand befindliches* zu repräsentieren. Was sich also im Moment des Luzidewerdens auf dramatische Weise verändert, ist das vom Gehirn aktivierte Selbstmodell.

Klarträumer können bestätigen, daß dieses Zu-sich-Kommen auch auf der Erlebnisebene sehr stark dem Sich-erinnern an längst vergessene und äußerst wichtige Informationen gleicht. Das darf als ein starkes Indiz für die Bedeutung der Gedächtnisfunktionen für das Zustandekommen von Luzidität gelten. Träumende informationsverarbeitende Systeme befinden sich in einem sehr seltsamen und für sie problematischen Zustand: Bei einer weitgehenden Input-Blockade (begleitet von einer Lähmung der Effektoren, die es ihnen vorübergehend unmöglich macht, diese Blockade durch reafferente Signale wieder aufzuheben; also: sich zu wecken) sind sie plötzlich mit einer internen Signalquelle konfrontiert.[105] Stehen dem System die normalen Lern- und Erinnerungsfunktionen nicht zur Verfügung, so entstehen aus dem verzweifelten Versuch, das weitgehend chaotische Signalgewitter zu einem internen Modell der Wirklichkeit zu verarbeiten, bizarre mentale Episoden – die nicht erinnert werden können. Ändern sich die physischen Randbedingungen, dann gewinnt das träumende System genau in dem Masse, in dem wieder gespeicherte Information verwendet werden kann, die Fähigkeit, diese inneren Ereignisse anhand vergangener Erlebnisse zu interpretieren und ein Selbstmodell zu erzeugen. Wenn es ein Selbstmodell gibt, werden manche der internen Zustände zu meinen Zuständen und die psychische Qualität subjektiven Erlebens tritt auf. Zustandsklarheit im Sinne von Luzidität erfordert eine zusätzliche Repräsentation der „inneren Geschichte" des Systems, die Fähigkeit, einen repräsentationalen Gesamtzustand im Lichte früherer repräsentationaler Gesamtzustände *als einen solchen* zu erkennen.[106] Genau diese metakognitive Leistung kann nur durch die Erzeugung eines stabilen Selbstmodells erbracht werden.

Ich hoffe, daß meine Leser durch diese knappe Schilderung einiger Fallbeispiele einen Eindruck davon gewonnen haben, wie abweichende Formen von mentaler Selbstmodellierung beim Menschen zu veränderten Zuständen phänomenalen Bewußtseins führen können. Manche der so entstehenden subjektiven Realitäten können aus der Perspektive des Erle-

[105] Wie ich bereits im vorangegangenen Kapitel gesagt habe, könnten dies die sogenannten PGO-Wellen aus dem Hirnstamm sein; vgl. Hobson 1987.

[106] Eine eng verwandte Theorie über das Entstehen luzider Träume vertritt Sue Blackmore, von der ich in persönlichen Gesprächen viel gelernt habe. Vgl. Blackmore 1982b, 1985, 1988.

benden Episoden des Entsetzens sein, andere als tiefe Bereicherungen des inneren Lebens erscheinen. Für alle diese Zustände gilt jedoch, daß sie einer naturalistischen Erklärung zugänglich sind, weil sie auf hochkomplexen Formen interner Informationsverarbeitung durch menschliche Gehirne beruhen, während – ähnlich wie bei den bereits geschilderten devianten mentalen Modellen der *Welt* – die vorwissenschaftlich-lebensweltliche Alltagspsychologie und die klassischen philosophischen Subjekttheorien an ihnen scheitern. Wenn dies richtig ist, dann muß es auch möglich sein, einen naturalistischen Begriff des Subjekts phänomenalen Erlebnisse zu entwickeln.

Deshalb habe ich in diesem Kapitel verschiedene Formen, in denen sich ein informationsverarbeitendes System intern selbst darstellen kann, analysiert. Ich habe darauf aufbauend den Begriff des Selbstmodells als eines inneren Analogrepräsentats eingeführt, der zu dem es konstruierenden System in einer mereologischen Repräsentationsbeziehung steht und es zu einem selbstähnlichen physischen System macht. Ich habe dann versucht, deutlich zu machen, daß dieser neue Begriff empirisch gut verankerbar ist und sich für die Beschreibung und Analyse unserer eigenen phänomenalen Zustände eignet – auch und gerade wenn es sich um abweichende phänomenale Zustände handelt. In dem nun folgenden letzten Abschnitt des Kapitels werde ich die zentrale These dieser Arbeit vorstellen und erläutern.

3.3 „SMT“: Die Selbstmodell-Theorie der Subjektivität

Natürliche Repräsentationssysteme instantiieren unter bestimmten Bedingungen psychologische Eigenschaften. Die höchste und komplexeste uns bekannte Eigenschaft dieser Art besteht in dem Besitz von Selbstbewußtsein. Selbstbewußtsein ermöglicht es einem System als Ganzem in eine besondere Beziehung zu einer Teilmenge seiner eigenen Zustände zu treten. Es gibt starke Indizien dafür, daß diese besondere Beziehung eine *Repräsentationsbeziehung* ist, die durch physisch realisierte Prozesse der Informationsverarbeitung innerhalb des betreffenden Systems hergestellt wird. Diese Prozesse erzeugen eine komplexe, physikalisch realisierte interne Datenstruktur – ein Selbstrepräsentat oder Selbstmodell. Bezüglich der fraglichen psychologischen Eigenschaft und dieses Selbstmodells werde ich nun eine These formulieren, die empirisch plausibel ist und eine naturalistische Perspektive auf das psychologische Subjekt innerer Zustände eröffnet. Ich werde diese These ab jetzt als *Selbstmodell-Theorie der Subjektivität* oder *Selbstmodell-Hypothese* – kurz: „**SMT**“ – bezeichnen.

> **(SMT):** Subjektivität ist eine Eigenschaft komplexer informationsverarbeitender Systeme, die genau dann instantiiert wird, wenn das System in das von ihm aktivierte Realitätsmodell ein Selbstmodell einbettet.

Zunächst einige erläuternde Bemerkungen:

– Mit „Subjektivität" ist hier der Kernaspekt psychologischer Subjektivität gemeint, den ich im ersten Kapitel dieses Buches als *relationale* Subjektivität bezeichnet habe. Sie verleiht manchen der vom System erzeugten inneren Zustände die phänomenale Qualität der „Meinigkeit". Diese Qualität wiederum besteht in nichts anderem, als daß die betreffenden mentalen Modelle in das Selbstmodell *eingebettet* sind. Subjektivität ist also letztlich eine Beziehung zwischen Datenstrukturen.

– Mit „Subjektivität" sind deshalb nicht der eingangs analysierte „Bewußtheitsaspekt" und der „Präsentationsaspekt" des spezifisch menschlichen phänomenalen Bewußtseins gemeint. Ich habe an anderer Stelle (Abschnitte 2.1.3 und 2.1.4) provisorische Hypothesen bezüglich einer Analyse dieser psychischen Phänomene im Rahmen einer naturalistischen Theorie mentaler Repräsentation formuliert. Im spezifisch humanen Wachbewußtsein fallen die drei Phänomene – unter Standardbedingungen – *faktisch* zusammen, von ihrer Analyse im Rahmen einer Theorie interner Repräsentation her sind sie allerdings nicht identisch und darum begrifflich disjunkt. Um das besser zu verstehen, was ich die Zentrierung repräsentationaler Gesamtzustände genannt habe, ist der Aspekt der Selbstmodellierung der ausschlaggebende. Die eben angesprochene Zentrierung erscheint auf der Ebene inneren Erlebens als die Perspektivität phänomenalen Bewußtseins und erzeugt die bekannten philosophischen Probleme.

– Das logische Subjekt der Prädikation ist immer *das System als Ganzes.* Das heißt: Subjektivität ist eine psychologische Eigenschaft, die einer bestimmten Klasse von informationsverarbeitenden Systemen zugeschrieben wird, nämlich genau solchen Systemen, die gerade ein Selbstmodell in ihr Realitätsmodell einbetten. Sprechen wir dabei von Systemen, denen wir den Status einer Person zubilligen[107], so ist Subjektivität eine Eigenschaft der personalen Beschreibungsebene. Personen besitzen die psychologische Eigenschaft Subjektivität genau dann, wenn sie Selbstmodellbesitzer sind.

– Alternativ dazu kann man „Subjektivität" auch von repräsentationalen Gesamtzuständen oder einzelnen mentalen Modellen prädizieren – das ist aber nicht der hier primär intendierte Sinn. Ein repräsentationaler Gesamtzustand ist dann ein subjektiver, wenn in ihm (genau[108]) ein Selbstmodell enthalten ist. Einzelne mentale Modelle sind dann subjektiv bzw. Inhalte subjektiven Bewußtseins, wenn sie in ein Selbstmodell eingebettet sind. Bei diesem erweiterten Gebrauch von „subjektiv" handelt es sich nicht um die Zuschreibung einer psychologischen Eigenschaft, weil solche Eigenschaften nur Personen oder Systemen als Ganzen zukommen. Eine solche Zuschreibung zeichnet die fraglichen internen Zustände jedoch als

[107] Ich gehe hier davon aus, daß der Begriff der Person eine letzte normative Komponente nicht verlieren wird. Vgl. Dennett 1981.

[108] Über multiperspektivische Bewußtseinszustände sagt **SMT** in der hier angebotenen Formulierung nichts aus. Ein Grund dafür ist, daß unser empirisches Wissen über multiple Selbstmodelle derzeit noch zu bruchstückhaft ist.

genau diejenigen Personenzustände oder Systemzustände aus, *aufgrund* deren Besitz wir eine Person oder ein System als psychologisches Subjekt charakterisieren.

– „Komplexe informationsverarbeitende Systeme" sind diejenige Klasse physischer Systeme, die prinzipiell – durch interne Selbstmodellierung – in der Lage sind, die fragliche psychologische Eigenschaft zu instantiieren. Komplex müssen diese Systeme sein, weil es mit Sicherheit starke, physikalisch oder biologisch zu beschreibende Randbedingungen *(„bottom-up constraints")* für die Erzeugung eines stabilen Selbstmodells gibt. Man muß davon ausgehen, daß die (noch zu leistende) Beschreibung des notwendigen Ausmaßes an physischer Komplexität sehr umfangreich sein wird. Das bedeutet nicht, daß diese Komplexität selbst in irgendeinem Sinne als Explanans für das Entstehen der fraglichen psychologischen Eigenschaft geltend gemacht werden soll.

– Sollte sich zeigen, daß Selbstbewußtsein ein „emergentes"[109] Phänomen ist, das in Systemen auch ohne interne Selbstmodellierung einfach ab einem gewissen kritischen Niveau struktureller oder funktionaler Komplexität auftritt, dann ist **SMT** falsifiziert. Dasselbe gilt, wenn es sich herausstellt, daß es nicht die Eigenschaften der Informationsverarbeitung und internen Repräsentation sind, die physische Systeme besitzen müssen, um Subjektivität zu instantiieren.[110] Sollten die für wissenschaftliche Erklärungen von Selbstbewußtsein interessanten Eigenschaften physikalischer Systeme einen *anderen* explanatorischen Set (eine andere Supervenienz-Basis-Familie) bilden als den hier immer schon vorausgesetzten, dann ist **SMT** ebenfalls falsifiziert.

– Die durch **SMT** vorgenommene Auszeichnung einer Klasse von Systemen ist indifferent gegenüber der Unterscheidung zwischen künstlichen und natürlichen[111] Systemen. Sollte sich zeigen lassen, daß prinzipiell nur biologische Systeme in einem interessanten Sinn Selbstmodellbesitzer sein können, dann können auch nur solche Systeme psychologische Subjektivität instantiieren. Die Entscheidung bezüglich dieser Frage hängt essentiell von der zugrunde gelegten Theorie der *Intentionalität* ab, speziell von unserer Antwort auf die Frage, wie stark phänomenaler Gehalt mit intentionalem Gehalt verknüpft ist. Ein anderer wichtiger Punkt ist in diesem

[109] Eine solche Entdeckung wäre erst dann möglich, wenn wir eine überzeugende wissenschaftstheoretische Konzeption emergenter Verursachung besäßen; vgl. Kapitel 2, Fußnote 111.

[110] Das betrifft die – falliblen – Hintergrundannahmen einer naturalistischen Theorie des Geistes als einer Theorie von *Repräsentationssystemen*, die ich in dieser Arbeit einfach vorausgesetzt habe. Natürlich könnten letztlich auch ganz andere Ansätze zielführend sein als der der Informatik und Repräsentationstheorie. Vgl. etwa Edelman 1987 für eine *populationstheoretische* Analyse der Genese höherer kognitiver Funktionen beim Menschen. Die These, daß Gehirne keine informationsverarbeitenden Systeme sind, vertritt Searle 1992: 222ff.

[111] Ich gehe hier der Einfachheit halber davon aus, daß die *künstlich/natürlich*–Unterscheidung exklusiv und erschöpfend ist. Neue Entwicklungen in der hybriden Biorobotik zeigen allerdings, das dies nicht der Fall ist.

Zusammenhang die Frage, was es heißt, daß ein System Interessen besitzt oder verfolgt – zum Beispiel, indem es abstrakte Instrumente wie mentale Modelle im allgemeinen oder Modelle des Selbst im besonderen entwickelt und einsetzt, um *Ziele* zu verfolgen. Ich habe keine Antworten auf diese beiden Fragen gegeben, deshalb können wesentliche Einschränkungen bezüglich der intendierten Klasse von informationsverarbeitenden Systemen notwendig sein. Prinzipiell schließt **SMT** jedoch die Möglichkeit künstlicher Subjektivität in keiner Weise aus.

– Die „Einbettung eines Selbstmodells in ein Realitätsmodell" soll die zentrale notwendige Bedingung für das Entstehen der psychologischen Eigenschaft „Subjektivität" in einem System sein. (Ich gehe dabei von der Notwendigkeit eines grundsätzlichen „In-der-Welt-Seins" aus, also davon, daß es subjektive Inhalte des Bewußtseins nur dann geben kann, wenn es auch objektive Inhalte gibt. Systeme, die nur ein Selbstmodell erzeugen ohne es in ein Weltmodell einzubinden, sind undenkbar: Ein Selbstmodell *ohne* Zentrierungsfunktion ist einfach nur ein nicht-zentriertes Realitätsmodell ohne phänomenale „Meinigkeit".) Ich habe dafür argumentiert, daß dies für unser eigenes phänomenales Bewußtsein eine empirisch plausible Annahme ist und schlage deshalb eine begriffliche Extrapolation vor. **SMT** ist damit auch der Versuch, einen neuen naturalisierten Begriff des Erlebnissubjekts anzubieten.

– Ob **SMT** mit interner Selbstmodellierung auch die *hinreichende* Bedingung für die Zuschreibung von Subjektivität angibt, ist eine offene empirische Frage. Und dies muß eines der zentralen Merkmale des naturalistischen Begriffs der Subjektivität sein: Er ist offen für zukünftige Anreicherungen, Präzisierungen und auch für eine *Elimination* durch die empirische Psychologie. **SMT** schlägt also im Sinne einer metatheoretisch-interdisziplinären Arbeitshypothese einen offenen Begriff einer bestimmten psychologischen Eigenschaft relativ zu einer Systemklasse vor.

– Um *bewußte* Subjektivität in Gestalt menschlichen Selbstbewußtseins zu erfassen, muß mindestens noch das Kriterium einer teilweisen Metamodellierung des Selbstmodells erfüllt sein – nur die von einer entsprechenden Funktion erfaßten Bereiche des Selbstmodells werden auch Inhalte phänomenalen Bewußtseins sein.[112] Mit Blick auf ein spezielles physisches System wie das menschliche Gehirn mag es außerdem sehr fraglich sein, ob es Selbstmodellierung ohne gleichzeitige Erzeugung mentaler *Präsentate* – also: Subjektivität ohne *Qualia* – leisten kann. Dies sind Beispiele für weitere Anreicherungen des Subjektivitätsbegriffs, die in bezug auf ein bestimmtes physisches System notwendig sein können. **SMT** will das begriffliche Fundament solcher Anreicherungen sein, nicht eine metaphysische These.

[112] Daß es tatsächlich sehr stabile, komplexe und funktional aktive Selbstmodelle ohne begleitendes Bewußtsein geben kann, zeigen uns gut belegte Phänomene wie das nächtliche Schlafwandeln von Menschen. Aus der Perspektive klassischer *a-priori*-Theorien des Subjekts mag „unbewußte Meinigkeit" auf den ersten Blick wie ein Widerspruch im Beiwort erscheinen. In Wirklichkeit sind solche Theorien einfach falsch, was die *empirische* Einheit der Apperzeption angeht.

– Eine der Schwierigkeiten für die Präzisierung von **SMT** liegt in der Tatsache begründet, daß die anvisierte psychologische Eigenschaft derart komplex ist, daß es keine Hoffnung auf auch nur annähernd vollständige und überzeugende Instantiierungserklärungen[113] für diese Eigenschaft in der nahen Zukunft gibt. Selbstmodellierung ist ein sehr umfassender Vorgang, der wahrscheinlich auf fast alle repräsentationalen Ressourcen des menschlichen Gehirns zugreift. Deshalb ist das mit seiner Aufklärung befaßte *empirische* Projekt ein wesentlich umfangreicheres als etwa eines, das nur eine gewisse eng umgrenzte Teilmenge mentaler Modelle (zum Beispiel Farbwahrnehmungen oder Gefühle) untersucht. Es ist ebenfalls klar, daß empirische Anreicherungen des Terms „Selbstmodell" diesen zum Gegenstand weitreichender begrifflicher Revisionen machen können – das ist ja gerade seine Stärke.

– Die Selbstmodell-Hypothese sagt nichts über logische, grammatische oder sozial konstituierte[114] Subjektivität, weil sie ausschließlich über eine psychologische Eigenschaft spricht. Allerdings geht mit ihr die These einher, daß für alle Zwecke einer theoretischen Psychologie das „Subjekt" oder das „Ich" als Erklärungsbegriff durch den des „Selbstmodells" substituiert werden kann: Alle problematischen Phänomene des Selbstbewußtseins können ohne einen Verlust an Erklärungs- und Beschreibungspotential auf der *repräsentationalen Ebene* aufgefangen werden. Psychologie ist somit *Selbstmodellforschung*; eine Wissenschaft, die sich mit dem repräsentationalen Gehalt, den funktionalen Eigenschaften und der physikalischen Realisierung einer besonderen Klasse von Datenstrukturen auseinanderzusetzen hat. Diese begriffliche Ansiedelung unseres Selbstbewußtseins auf der Ebene einer „phänomenalen Informatik" soll eine neue Beschreibungsebene öffnen, aber nicht eine neurobiologische Reduktion unseres Innenlebens vorbereiten. Im Gegenteil: Eine empirisch verankerbare naturalistische Terminologie erweitert die Möglichkeiten menschlicher Selbsterkenntnis und Selbstbeschreibung. Denn der Begriff des „Selbstmodells" besitzt keine metaphysischen Konnotationen und vermeidet all jene notorischen Äquivokationen, die allein dadurch entstehen, daß der metaphysische Subjektbegriff in der Geistesgeschichte des Abendlandes eine so ungemein dominante Rolle gespielt hat. Es wird sich zeigen müssen, welche philosophische Potenz die nicht-psychologischen Elemente des klassischen Subjektbegriffs besitzen, wenn eine Verquickung der traditionellen semantischen Elemente mit der Frage phänomenalen Selbstbewußtseins nicht mehr gelingt: Kann das transzendentale Subjekt ohne das empirische Ich überleben?[115]

Soweit die ersten Erläuterungen zu **SMT** und ihrer Rolle in einer naturalistischen Theorie des Geistes. In den nächsten beiden Kapiteln werde ich

[113] Vgl. Cummins 1983, Eimer 1990.

[114] Allerdings haben eine Reihe von Autoren vermutet, daß Selbstmodelle durch die repräsentationale Internalisierung sozialer Aktivitäten entstehen. Vgl. hierzu Oatley 1988.

[115] Vgl. dazu auch Horstmann 1990.

auf der Grundlage dieser Hypothese und der Überlegungen, die uns in den vorangegangenen Kapiteln zu ihr geführt haben, Antworten auf die am Ende des ersten Kapitels formulierten Fragen zu geben versuchen. Diese Antworten werden zu einer weiteren Verdeutlichung des Gedankens führen. Im weiteren Verlauf werde ich nun auch einige Einwände diskutieren, um auf diese Weise noch verbliebene Unklarheiten auszuräumen. Zunächst jedoch möchte ich prominente Beiträge aus der analytischen Gegenwartsphilosophie einsetzen, um zusätzliche Perspektiven auf die philosophische Problematik zu eröffnen und unser Bild der durch sie gebildeten theoretischen Landschaft ein weiteres Mal anzureichern und zu differenzieren.

4. Kapitel

Fledermäuse, objektive Selbste und die Irreduzibilität der Innenperspektive

4.1 *Inneres Wissen:* Thomas Nagels Philosophie der Subjektivität

Es ist ein Verdienst Thomas Nagels, das Problem der Subjektivität des Mentalen in einem eher ungünstigen intellektuellen Klima auf originelle Weise neu formuliert und behandelt zu haben. Er hat dies in einer Diskussionsphase der analytischen Philosophie des Geistes getan, in der starke Tendenzen zur Ausblendung des gesamten Problemkomplexes überhaupt bestanden. Es waren dies allerdings Tendenzen, die ihren Ursprung in Grundannahmen über die Natur mentaler Phänomene haben, welche der Nagelschen Position zumindest eng verwandt sind. Seine intellektuelle Sensibilität gegenüber derjenigen Eigenheit psychischer Zustände, die wir als ihre Subjektivität zu bezeichnen pflegen, und den mit ihr verbundenen theoretischen Schwierigkeiten zeigt sich schon in einer seiner frühen Arbeiten zum Leib-Seele-Problem.

Bereits in seinem 1965 veröffentlichten Artikel *„Physicalism"*[1] untersuchte Nagel die Gründe für sein Abgestoßensein vom Physikalismus als einer Theorie der Relation mentaler und physischer Zustände, obwohl er von der Wahrheit dieser Theorie überzeugt war. Selbst nachdem alle Einwände gegen eine materialistische Theorie der Identität zwischen physischen und mentalen Phänomenen (etwa solche, die die Privatheit oder den epistemologisch privilegierten Zugang zum Bereich des Mentalen thematisieren) widerlegt sind, bleiben Zweifel: Das Subjekt psychologischer Eigenschaften scheint niemals in einem Körper oder einer sonstigen objektivwissenschaftlich zu beschreibenden Substanz aufgehen zu können. Denn, so schrieb Nagel schon damals[2], das Subjekt psychologischer Attribute ist ein *Selbst*; nirgendwo in einer noch so detaillierten Beschreibung meines Körpers oder meines Gehirns ist Raum für die Tatsache, daß die eng mit seinen Zuständen korrelierten oder sogar durch es erzeugten mentalen Zustände *meine* Zustände sind. Sie gehören in einem *nicht-derivativen* Sinne zu ihrem Subjekt – und das ist niemals ein innerweltlicher Gegenstand, sondern immer ein Selbst oder ein Ich. Aber auch nicht-physische Substanzen eignen sich, so Nagel, nicht als Kandidaten für diese Leerstelle in unserer Theorie des Geistes, weil nichts an ihrer Bestimmung darauf hin-

[1] Vgl. Nagel 1965: 339–356.

[2] Vgl. Nagel 1965 und die deutsche Übersetzung des Originalaufsatzes in Bieri 1981: 66ff.

weist, daß sie meine sind. Die Relation zwischen dem Subjekt und seinen psychischen Zuständen kann darum nicht nach dem Modell einer Substanz-Attribut-Relation analysiert werden, da wahre psychologische Subjektivität eine vollkommen andere logische Struktur zu besitzen scheint. Die Beziehung zwischen mir als Erlebendem und meinen inneren Zuständen hat ein Moment der Essentialität, der Innerlichkeit und Gewißheit, das durch Denkfiguren wie die oben angesprochene deskriptiv nicht aufgefangen wird.

Ihren vorläufigen Abschluß haben Thomas Nagels Analysen zum Problem der Perspektivität mentaler Zustände in seinem Aufsatz *„The Objective Self"*[3] gefunden. Der Hauptzweck des nun folgenden Kapitels liegt in einer knappen Rekonstruktion seiner Gedankengänge, die dann als Grundlage einer Kritik seiner Position dienen soll. Im nächsten Kapitel werde ich dann abschließend versuchen, die sich aus dieser Kritik ergebenden Fragen in einer Zusammenführung mit den bereits zu Beginn formulierten systematischen Fragen zu beantworten. Thomas Nagels Beiträge scheinen paradigmatisch geeignet für ein weiteres Eindringen in die Problematik, da sie auf nicht-technische Weise und doch in konzentrierter Form eine große Anzahl wesentlicher Facetten des Subjektivitäts-Problems beleuchten – eines Problems, das seit einiger Zeit kein exklusiv philosophisches Problem mehr ist, sondern auf der Grenze zwischen Philosophie und teilweise sehr jungen Einzelwissenschaften angesiedelt ist. Wie wir sehen werden, bedingt allein diese Verschiebung der Problemposition bereits die Notwendigkeit grundsätzlich neuer Lösungsstrategien.

4.2.1 „Wie es ist, ein X zu sein"

Im Jahre 1974 veröffentlichte Thomas Nagel einen Aufsatz unter dem Titel *„What is it like to be a bat?"*[4], der bald die Aufmerksamkeit vieler Philosophen auf sich zog und eine große Anzahl von Erwiderungen und Kritiken hervorrief.[5] Dieser kleine Beitrag ist mittlerweile zu einem modernen Klassiker der Philosophie des Geistes avanciert. Ein Grund für die andauernde Popularität des kurzen Artikels mag sein, daß er eine neue, intuitiv leicht zugängliche Formulierung dafür anbietet, was es heißt, daß manche mentalen Zustände neben intentionalem Gehalt auch noch einen *phänomenalen*

[3] In Nagel 1986, vgl. die deutsche Übersetzung von Michael Gebauer in Nagel 1992; eine frühe und leicht veränderte deutsche Version findet sich in Siep 1983.

[4] *The Philosophical Review* **83** (1974); wiederabgedruckt in *Mortal Questions*, Cambridge 1979, deutsch in Bieri 1981 und Nagel 1984.

[5] Eine knappe Auswahl der wesentlichsten Texte würde Armstrong/Malcolm 1984, Churchland, P.M. 1985a, Churchland, P.S. 1986, Dennett 1988, Dennett/Hofstadter 1987(1981), Foss 1987, Jackson 1982, Lewis 1984, Lycan 1987 (Kapitel 7), Lycan 1990, Maloney 1985, McGinn 1983, McMullen 1985, Nemirow 1979 (vgl. auch 1980), Peacocke 1989, Tye 1986, Van Gulick 1985 beinhalten müssen.

oder qualitativen Gehalt besitzen. Die neue Formel für diese Tatsache, die Nagel uns anbietet, lautet: *Es ist irgendwie*[6], das Subjekt dieser mentalen Zustände zu sein. Das heißt: Für jedes System mit internen Zuständen, die einen phänomenalen Gehalt besitzen, scheint es eine private Qualität zu geben, die eines der interessantesten Merkmale dieser inneren Zustände ist. Diese Qualität besteht darin, daß es *irgendwie ist*, sich in diesen Zuständen zu befinden – beziehungsweise: dieses System zu *sein*.

Intuitiv erscheint diese Charakterisierung mentaler Zustände über ihren phänomenalen Gehalt augenblicklich einleuchtend. Denn wenn wir uns fragen, ob ein bestimmtes biologisches oder künstliches System ein mentales Innenleben besitzt, meinen wir immer genau dies: *Fühlt es sich irgendwie an*, dieses System zu sein? Gibt es in diesem System eine Instanz, die einen *unvermittelten* Zugang zu einer Teilmenge seiner internen Zustände hat? Gibt es eine Instanz, die auf direkte Art und Weise erlebt, wie es ist, eben jenes System zu sein? „Wie es ist" faßt den qualitativen Gehalt aller aktuellen, bewußten mentalen Zustände eines Systems zusammen und nimmt dabei bezug auf eine vorsprachliche Gegebenheit unseres psychischen Lebens, die jedem von uns zuerst einmal prätheoretisch einsichtig ist. Uns allen ist gemeinsam, daß es *irgendwie für uns ist*, wir selbst zu sein. Wie immer eine zukünftige Theorie des Geistes aussehen mag, sie wird nichts daran ändern, daß es *irgendwie ist*, sich in den von ihr auf eine neue Weise beschriebenen geistigen Zuständen zu befinden – und jede zukünftige Theorie des Geistes wird eine gute Erklärung für das Zustandekommen dieses zentralen Charakteristikums mentaler Phänomene anbieten müssen, wenn sie uns überzeugen will.

Betrachtet man diese erste rudimentäre Entfaltung des Problems, so sieht man, wie wenig Nagels Angebot des „Wie es ist, ein X zu sein" bei einer Beantwortung der zentralen Fragen hilft: In der Analyse führt es nicht weiter. Es leistet aber die deutliche Benennung des psychologischen Aspekts eines Problems, über das sich seit Descartes viele Philosophen den

[6] Um philosophiegeschichtlich genau zu sein, muß man darauf hinweisen, daß diese Formulierung zuerst von B.A. Farrell eingesetzt wurde. Farrell diskutierte in einem Aufsatz in der Zeitschrift *Mind* nicht nur die *„raw feels"* von Marsmenschen und Robotern, sondern benutzt auch bereits das erst durch Nagel berühmt gewordene Beispiel der Fledermaus (vgl. Farrell 1950: 183). Interessant an seinen sprachanalytischen Untersuchungen des Begriffs *„experience"* ist zudem die Tatsache, daß in ihnen bereits klassische Argumentationsfiguren derjenigen Variante der Identitätstheorie entwickelt werden, die später als „eliminativer Materialismus" einen großen Bekanntheitsgrad erreichte. Auch hier zeigt sich eine enge Verwandschaft mit späteren Positionen (etwa von Rorty oder Churchland), die bis in die verwendeten Beispiele hinein reicht: *„If the relevant sciences go on developing in this direction, and if Western societies assimilate their work, than [sic] it is quite possible that the notion of „experience" will be generally discarded as delusive. If and when this happens, our present philosophical difficulties about it will disappear. But it is just because we are in conflict, and perhaps transition, about the notion that we cannot either accept or reject it at present without absurdity or falsehood. In these respects, the notion of „experience" can be shown to resemble an occult notion like „witchcraft" in a primitive community that is in the process of being acculturated to the West."* (Farrell 1950: 195)

Kopf zerbrochen haben. Es hat zudem den Vorteil, dieses Ziel zu erreichen, ohne metaphysische Implikationen oder die Hypostasierung der fraglichen Eigenschaft zu erzwingen. Wichtig an Thomas Nagels Formulierung von subjektivem Gehalt ist deshalb ihre ontologische Neutralität: Wenn wir konzedieren, daß es mit der Innenperspektive faktisch eine wichtige Dimension der psychischen Welt gibt, die wir noch nicht verstanden haben, zwingt uns dies nicht zu ontologischen Schlüssen – ganz abgesehen davon, daß es ein Markenzeichen unseriösen Philosophierens ist, immer dort, wo wir ein Phänomen nicht verstanden haben, eine metaphysische Entität zu postulieren und auf diese Weise Unwissenheit in Gespenster zu transformieren. Man kann zugeben, daß es das Nagelsche „Wie es ist, ein X zu sein" gibt, auch wenn man noch nicht erklären kann, wie es *in dieser Welt* sein kann. Und schließlich wirft Nagel folgerichtig auch die Frage auf, welche Methode der Erkenntnis uns überhaupt zur Verfügung steht, um uns der Subjektivität unserer eigenen inneren Zustände anzunähern. Diese Frage wiederum lenkt unser Augenmerk auf die Notwendigkeit einer möglichst genauen Beschreibung des Explanandums zurück. Nagel schreibt:

> Die Tatsache, daß ein Organismus *überhaupt* bewußte Erfahrung hat, heißt im wesentlichen, daß es irgendwie ist, dieser Organismus zu *sein*. Es mag weitere Implikationen bezüglich der Form der Erfahrung geben; es mag sogar (obwohl ich es bezweifle) Implikationen bezüglich des Verhaltens des Organismus geben. Grundsätzlich aber hat ein Organismus bewußte mentale Zustände dann und nur dann, wenn es irgendwie ist, dieser Organismus zu *sein* – wenn es irgendwie *für* diesen Organismus ist.[7]

Die Qualität des *pour soi*, die Nagel hier zu einer notwendigen Bedingung für die Zuschreibung bewußter mentaler Zustände macht, erscheint als eine essentiell *private* Eigenschaft: Es sind immer nur *meine* mentalen Zustände, von denen ich berichten kann, wie es ist, sich in ihnen zu befinden. Eine der philosophisch interessantesten Fragen in diesem Zusammenhang ist die nach dem erkenntnistheoretischen Status von Selbstzuschreibungen psychologischer Eigenschaften, die ausschließlich introspektiv gewonnen werden: Ist das Subjekt epistemologisch privilegiert gegenüber seinen eigenen inneren Zuständen? Kann man auf klare und sinnvolle Weise behaupten, daß ich *weiß*, wie es jetzt gerade ist, *ich* zu sein? Introspektion ist die große Versuchung für alle Philosophen des Geistes. Ihre Gefährlichkeit resultiert aus der Tatsache, daß sie zumeist mit einem naiven Realismus bezüglich ihrer Objekte einhergeht und daß strukturell fixierte Grundannahmen unseres zentralen Nervensystems über das Wesen der Wirklichkeit zwangsläufig und unmerklich in die von ihr erzeugten höherstufigen Formen mentalen Gehalts einfließen – denn Introspektion ist, wie ich bereits zu erläutern versucht habe, eine Variante desjenigen Vorgangs, in dem unser aller cartesianische Intuitionen wurzeln. Andererseits kann die Beschäftigung mit der Philosophie des Geistes gerade deshalb besonders interessant sein, weil sie

[7] Vgl. Bieri 1981: 262.

sowohl den Vorgang der Introspektion wie auch die innere Wirklichkeit des Philosophen schrittweise und unausbleiblich verändert. Und was gibt es Schlimmeres als Philosophie, durch die wir uns nicht mehr verändern?

Die Anerkennung der mentalen Qualität „Subjektivität" führt sämtliche seit Ryle entwickelten Formen des Physikalismus und des Funktionalismus in Aporien. Für die Identitätstheorie läßt sich nicht mehr zeigen, wie *mein* Besitzen eines bestimmten psychologischen Attributs identisch sein soll mit dem Besitzen eines solchen Attributs durch eine wie auch immer naturwissenschaftlich zu beschreibende Substanz.[8] Die Eliminationsvarianten[9] der Theorie kommen in größte Schwierigkeiten, was den Begriff der Wissenschaft als eines von rationalen Subjekten betriebenen Unternehmens angeht und funktionale Analysen sind notorisch resistent gegen Qualia und Subjektivität: Etwas könnte eine perfekte funktionale Simulation von mir oder einem meiner Leser sein, ohne daß es *irgendwie wäre*, dieses System zu sein. Denn es könnte ein System geben, das sich in genau denselben Input-Output-Relationen befindet wie ich, und das – auch was die interne Vernetzung funktional spezifizierter Ereignisse angeht – mit mir identisch ist. Trotzdem wäre es möglich, daß ein solcher – funktional isomorpher – Doppelgänger keine *Erlebnisse* hat.[10] Diese Resistenz der subjektiven Erlebnisperspektive gegen monistisch motivierte Analysen macht sie philosophisch interessant und man muß es als Verdienst Nagels werten, sie als wichtige Problemvariante in der neueren Philosophie des Geistes etabliert zu haben.

Welche Methoden stehen uns zur Verfügung, um zu entscheiden, ob das hier diskutierte Kriterium für die Zuschreibung von Bewußtsein für ein beliebiges System erfüllt ist? Der erste Kandidat scheint eine Art „empathischer Imagination" zu sein: Wir könnten versuchen, uns auf der Ebene der Vorstellung in eine Fledermaus, einen Marsmenschen oder einen Großrechner *einzufühlen*. Einer solchen „phänomenalen Emulation" sind durch unsere eigene biologische Struktur jedoch enge Grenzen gesetzt, nämlich in Gestalt der funktionalen Architektur unserer Nervensysteme, die nur sehr stark eingeschränkte Quellen und Mechanismen der Verarbeitung von Information zur Verfügung stellt um die psychische Simulation eines anderen geistigen Wesens durchzuführen. Über Phantasie und Vorstellung als

[8] Vgl. Nagel in Bieri 1981: 68.

[9] Die kanonischen Texte sind Churchland 1979, 1981, Feyerabend 1981(1970), Rorty 1981a(1965), 1981b(1970), 1981c(1979). Weitere Literaturangaben zur Entwicklung des eliminativen Materialismus finden sich in Bieri 1981 und Lycan 1990.

[10] Die zwei wichtigsten Versuche, die Intuition zu explizieren, daß es für jedes System mit physikalischen, funktionalen und phänomenalen Eigenschaften einen „bewußtlosen Doppelgänger" geben könnte, der dieselben physischen und funktionalen Merkmale besitzt *ohne* phänomenale Eigenschaften zu instantiieren, sind das *Absent-Qualia-Argument* (Block 1980, Block/Fodor 1972; dazu auch Churchland 1981 und Dennett 1979, 1981, Shoemaker 1981c) und das *Modal Argument* (Campbell 1970, Kirk 1974; die Repliken von Locke 1976, Lycan 1979, Shoemaker 1975; ferner Churchland 1981, 1985, Jackson 1982, Maloney 1985).

Instrumente zur Beantwortung der Frage nach dem Bewußtsein – etwa von Fledermäusen – schreibt Nagel:

> Insoweit ich mir dies vorstellen kann (was nicht sehr weit ist), sagt es mir nur, wie es für *mich* wäre, mich so zu verhalten, wie sich eine Fledermaus verhält. Das aber ist nicht die Frage. Ich möchte wissen, wie es für eine *Fledermaus* ist, eine Fledermaus zu sein. Wenn ich mir jedoch dies nur vorzustellen versuche, bin ich auf die Ressourcen meines eigenen Bewußtseins eingeschränkt, und diese Ressourcen sind für das Vorhaben unzulänglich.[11]

Objektive Methoden zur Feststellung des Vorhandenseins einer subjektiven Erlebnisperspektive innerhalb eines bestimmten Systems scheinen erst recht zum Scheitern verurteilt zu sein. Je objektiver die Aussagen werden, die wir über die internen Zustände von Personen machen, desto mehr entfernen wir uns von der subjektiven Perspektive desjenigen, dessen innere Erlebnisse diese Zustände sind. Wenn das aber wahr ist, dann scheint dies ein fundamentales Problem für jeden wissenschaftlichen Realismus in der Philosophie des Geistes im besonderen und in der *Cognitive Science* im allgemeinen zu sein: Wenn wir mentale Phänomene ernst nehmen wollen, dann müssen wir auch zugeben, daß ihr phänomenaler Gehalt eines ihrer essentiellen Hauptmerkmale ist. Wenn es sich nun aber zeigen sollte, daß sich dieser Gehalt wissenschaftlichen Analysen aus prinzipiellen Gründen entzieht, wenn wir also nicht mehr erklären können, wie mentale Phänomene in ihrem vollen Gehalt Teil der Welt sind, dann muß das Projekt einer naturalistischen Theorie des Mentalen bereits im Vorfeld als gescheitert gelten.

Für eine naturalistische Theorie des Mentalen könnten mentale Ereignisse in einer ersten Annäherung eine problematische und weitgehend unverstandene Klasse komplexer physikalischer Ereignisse sein, die sich dadurch auszeichnen, daß sie

a) **repräsentationalen Gehalt** besitzen; d. h. *für* ein bestimmtes System Teile der Welt oder seiner selbst repräsentieren, und

b) **kausale Rollen** einnehmen können, d. h. als *funktional* beschriebene Zustände von zentraler Bedeutung in bezug auf die Genese und/oder Erklärung von Verhalten sind, sowie

c) **phänomenalen Gehalt** aufweisen; d. h. im Bewußtsein spezifische Erlebnisqualitäten erzeugen, die essentiell an eine subjektive Perspektive gebunden zu sein scheinen.

Wie mentale Repräsentate ihre Referentialität erlangen und so manche Systeme zu intentionalen Systemen[12] machen, war vielleicht die zentrale

[11] Vgl. Bieri 1981: 264.

[12] Vgl. Dennett 1981. Daniel Dennett umkreist seit 1969 das Problem, was es bedeutet, einem System intentionale Zustände zuzuschreiben. Dabei hat er eine einflußreiche instrumentalistische Theorie solcher Zustände entwickelt, die davon ausgeht, daß wir bei bestimmten Systemen erfolgreiche Verhaltensprognosen nur noch dann durchführen können, wenn wir ihnen gegenüber eine bestimmte Einstellung einnehmen: die *intentional stance*. Diese intentionale Einstellung zeichnet sich durch Vorteile bei der Erklärung und Vorhersage komplexer

Frage der philosophischen Debatte der letzten Jahre. Mit Blick auf den zweiten Komplex von Problemen und auf die Diskussion des Leib-Seele-Problems seit dem Zweiten Weltkrieg stellt man ebenfalls fest, daß eine Fülle neuer Lösungsansätze auf einem hohen analytischen Niveau entwickelt worden ist.[13] Für die den ersten beiden eng verwandte und doch ganz andere Frage, unter welchen Bedingungen wir andere Systeme als zu unserer Klasse *bewußter* Wesen gehörig ansehen und sozusagen in die „Gemeinschaft geistiger Wesen" aufnehmen, scheint es dagegen unerläßlich, zunächst zu einer befriedigenden Theorie des qualitativen Gehalts, der Bewußtheit und der Subjektzentriertheit mentaler Repräsentate zu gelangen. Das Fledermaus-Argument verknüpft nun den ersten mit dem dritten Punkt: Die Nagelsche Qualität des „Wie es ist, ein X zu sein" stellt sich als eng verbunden mit einem phänomenalen *Identitätserlebnis* dar, das seine Wurzeln in der zuletzt genannten Subjektzentriertheit hat.

Denn wir sehen weiterhin, daß es neben dem angesprochenen Charakteristikum mentaler Zustände *(„Wie es ist . . .")* noch ein weiteres gibt, das mit Blick auf eine begriffliche Analyse und ihren wissenschaftstheoretischen Status große Bedeutung besitzt, nämlich ihre „Meinigkeit". In den üblichen Bewußtseinszuständen sind innere Prozesse immer meine inneren Prozesse: Es scheint eine Entität zu geben, die in einer Relation des Besitzens zu ihnen steht.[14] „Meinigkeit" ist die phänomenologische Formulierung für eine der wichtigsten Eigenschaften mentaler Zustände aus der Perspektive der ersten Person. Sie wird ein wichtiges Element jeder Theorie des psychologischen Subjekts sein, weil sie die Relationalität dieses Sub-

---

Verhaltensmuster aus, muß aber durch immer bessere physikalisch-funktionale Zusatzerklärungen supplementiert werden. Die Theorie intentionaler Systeme erkennt die Irreduzibilität der personalen Beschreibungsebene an, macht aber keine ontologischen Aussagen darüber, *was es gibt*. Vielmehr zeichnet sie eine Klasse von Systemen aus, bei der vorläufig nur ein bestimmter Erklärungstyp funktioniert und analysiert diesen Typ von Erklärung. Vgl. hierzu Dennett 1969, 1978, Beckermann 1986c und Bieri 1987c. Die wichtigsten Arbeiten Dennetts zu diesem Thema sind in dem Sammelband *„The Intentional Stance"* zusammengefaßt, vgl. Dennett 1987b.

[13] Seit dem Zusammenbruch des Ryleschen Projekts sind im wesentlichen neun Leib-Seele-Theorien entwickelt worden. Die wichtigsten Probleme der Debatte sind die nomologische Inkommensurabilität des Mentalen und die Frage nach der Theorieneutralität präreflexiver „innerer Gegebenheiten". Vgl. Metzinger 1985, 1990, 1991.

[14] Strawson hat diese Relation unter dem Aspekt der logischen Primitivität des Person-Begriffs näher untersucht. Er kommt dabei zu dem Ergebnis, daß wir – wenn wir mit dem „Haben" psychologischer Eigenschaften nicht die kontingente Tatsache ihrer kausalen Abhängigkeit von einem bestimmten Körper in der Welt meinen – das Erlebnissubjekt nicht ohne Rekurs auf öffentliche Personeneigenschaften identifizieren können. Wenn man dagegen wie Wittgenstein in einer bestimmten Phase oder Moritz Schlick eine quasi-buddhistische *„No-ownership"*-Theorie bezüglich der Beziehung zwischen Subjekt und internen Zuständen vertritt, erkauft man die sprachanalytische Auflösung des Problems mit deskriptiver Unplausibilität. Eine naturalistische Theorie der Subjektgebundenheit mentaler Zustände dagegen wird vor aller begrifflichen Analyse verständlich machen müssen, wie es überhaupt zum Auftreten der fraglichen psychologischen Eigenschaft in biologischen Gehirnen kommen konnte. Vgl. Strawson 1959.

jekts und seiner Zustände ins Spiel bringt. Man kann jedoch nicht sagen, daß „Meinigkeit" *a priori* allen mentalen Zuständen zukommt und sie so vor aller Erfahrung bereits begrifflich determiniert. „Meinigkeit" kann kontingenterweise verloren gehen – zum Beispiel, wenn ich einen schizophrenen Schub erlebe und mir andere Menschen ihre Gedanken senden oder plötzlich fremde mentale Zustände scheinbar von außen in meinen Bewußtseinsraum hineinforciert werden. Wir haben bereits gesehen, daß das menschliche Gehirn faktisch kognitive Zustände erzeugen kann, die auf der Ebene phänomenalen Bewußtseins als *fremdsubjektiv* (fremde Gedanken) oder als *nicht-subjektiv* (Stimmenhören) erlebt werden. Die Tatsache, daß die fragliche Eigenschaft mentaler Zustände auch in Abstufungen auftritt, verlorengehen kann und ganz offensichtlich stark durch Eigenschaften des Informationsflusses im Gehirn „von unten" determiniert ist, bildet überhaupt erst die Grundlage für das Unternehmen einer Naturalisierung des psychologischen Subjekts. Erst die reichhaltigen empirischen Erkenntnisse der letzten Jahrzehnte – etwa durch Untersuchungen an *split-brain*-Patienten oder solchen, die unter *Dissociative Identity Disorder* leiden – haben der Vermutung ihre Plausibilität verliehen, Subjektivität könnte eine natürliche Eigenschaft sein.

Das, was ich unter dem Namen „Meinigkeit" als eine Eigenschaft psychischer Zustände in Anschlag gebracht habe, ist eines der wichtigsten Elemente der phänomenalen Realität. „Meinigkeit" scheint sehr eng mit dem verknüpft zu sein, was Nagel als *„Wie es ist. . ."* bezeichnet: Diejenigen Zustände, von denen ich sagen kann *wie es ist*, sich in ihnen zu befinden, sind fast immer *meine*. Meist sind sie es in einem nicht-derivativen Sinne.[15] Das heißt: Ich als mich selbst erlebendes geistiges Wesen erarbeite oder produziere diese Eigenschaft meiner mentalen Zustände nicht – sie scheinen mir von Anfang an als meine *gegeben*[16] zu sein. Dies ist auch der Grund dafür, daß sich eine Substanz-Attribut-Relation nicht für die Analyse dieser Situation anbietet, weil sie eine Zufälligkeit zwischen Eigenschaften und ihren Trägern ins Spiel bringt, die kontraintuitiv ist, weil sie in nicht-pathologischen bzw. nicht-devianten Bewußtseinszuständen nicht erlebt wird.[17] Zunächst können wir festhalten, daß „Meinigkeit" eine relationale Eigenschaft bestimmter mentaler Repräsentate oder mentaler Modelle ist. Bis jetzt hat unser Blick auf Nagels frühe Arbeiten also die folgenden Aspekte der problematischen Eigenart mentaler Zustände, die wir häufig als ihre Subjektivität zusammenzufassen pflegen, berührt:

– **Phänomenaler Gehalt**: *Wie es ist*, sich in ihnen zu befinden;

---

[15] Vgl. Bieri 1981: 68.

[16] Den Mythos des nicht-inferentiellen, qualitativen Wissens durch „selbstbeglaubigende, non-verbale innere Episoden" und unmittelbare Gegebenheiten des Bewußtseins hat Wilfrid Sellars attackiert und aufzulösen versucht. Vgl. hierzu Sellars 1963, 1981 und Kurthen 1990: 21ff.

[17] Nagel hat bereits 1965 ausführlich auf diesen Punkt hingewiesen. Nicht alle Formen von Ich-Auflösung beim Menschen müssen Ich-Störungen im pathologischen Sinne sein. Vgl. etwa Dittrich 1985.

- **Relationalität**: Mentale Zustände sind *meine*, bzw. solche, von denen wir sagen, daß Personen oder Subjekte sie „besitzen";
- **Privatheit**: Mentale Zustände sind aus erkenntnistheoretischen Gründen problematisch, da sie nicht-öffentliche Eigenschaften exemplifizieren.

Man kann Nagel allerdings vorwerfen, daß er systematisch unser aller cartesianische Intuitionen ausbeutet. Diese sind zu einem nicht unbeträchtlichen Teil auch in Gestalt der abendländischen Alltagspsychologie fixiert und fließen deshalb leicht in theoretische Argumentationen ein, sofern diese nicht in formalen Kunstsprachen durchgeführt werden. Die erste und wichtigste dieser cartesianischen Intuitionen ist vielleicht die „Selbsttransparenz des Bewußtseins" *(„. . . Denn da ich jetzt weiß, daß ja selbst die Körper nicht eigentlich durch die Sinne oder durch die Fähigkeit der Einbildung, sondern einzig und allein durch den Verstand erfaßt werden, auch nicht dadurch, daß man sie betastet oder sieht, sondern, daß man sie denkt: so erkenne ich ganz offenbar, daß ich nichts leichter und augenscheinlicher erfassen kann – als meinen Geist.* "[18]). Diese Intuition läßt die meisten von uns glauben, daß wir uns über unser eigenes Bewußtsein nicht täuschen können: Im Bereich des Mentalen fallen Sein und Schein zusammen, *esse est experiri*.[19] Alle unsere Aussagen über das problematische Explanandum, die fragliche Subjektivität mentaler Zustände, sind – das darf in diesem Zusammenhang nicht übersehen werden – durch die Introspektion je einzelner gewonnen und beziehen sich also nicht auf öffentliche Ereignisse oder Eigenschaften. In diesem Sinne sind introspektive Aussagen zunächst nicht-überprüfbare Aussagen, wiewohl ihr heuristischer Wert für eine rationale Problemexposition nicht in Zweifel stehen kann. Es fragt sich also, welchen erkenntnistheoretischen Status wir dem Vorgang der Introspektion zuweisen. Ich plädiere dafür, ihn nicht als eine theoretisch relevante Form von „Wissen" zu interpretieren[20] ohne gleichzeitig in neobehavioristische Radikalismen zu verfallen. Dann jedoch stellt sich die Frage, wie man sich der Subjektivität unserer Innenwelt *überhaupt* annähern kann: Denn schließlich führen uns objektive Analysen immer weiter weg von der eigentümlichen Konkretheit und Individuenbezogenheit des Phänomens. Es sieht vorerst so aus, als könne man sich dem Problem der Perspektivität mentaler Zustände weder durch das Einnehmen noch durch das schrittweise Eliminieren von Perspektiven nähern.

Den Wissensaspekt von subjektiven Zuständen, insbesondere von Qualia, hat Frank Jackson bei der Entwicklung seines *Knowledge-Arguments* in den Mittelpunkt gerückt.[21] Ausgangspunkt des Arguments sind zwei Ge-

[18] Vgl. Descartes 1972 (1641): 26.

[19] Vgl. Bieri 1981: 206.

[20] „Theoretisch relevant" heißt in diesem Zusammenhang: „Nicht in der Lage, Wahrheitsansprüche im Sinne des wissenschaftlichen Erkenntnismodells zu stellen". Ich beziehe mich hier auf den Unterschied zwischen digitalem Wissen und analogem Wissen, den ich in Abschnitt 2.2.1 näher erläutert habe.

[21] Vgl. Jackson 1982; die drei Erwiderungen von Churchland 1985, die Antwort von Jackson 1986 sowie die entsprechende Replik von Churchland 1989. Aus einer

dankenexperimente, die deutlich machen sollen, daß unser physikalisches Wissen über Körperempfindungen oder Wahrnehmungserlebnisse unvollständig sein muß, weil es einen zweiten Typ von „phänomenaler Information" gibt, den wir scheinbar nur durch *qualitatives Wissen* erwerben können.[22] Im ersten Gedankenexperiment begegnen wir Fred. Fred besitzt ein weit überdurchschnittlich entwickeltes Farbunterscheidungsvermögen, wie wissenschaftliche Untersuchungen zeigen – noch nie hatte jemand eine derartig differenzierte Form von visueller Farbwahrnehmung. Fred kann einen Haufen reifer Tomaten in zwei Haufen sortieren, weil er zwei Farben – „Rot_1" und „Rot_2" – sieht, wo *wir* nur eine sehen. Wenn man ihm die Augen verbindet, die Tomaten wieder vermischt und ihn den so entstandenen Haufen ein zweites Mal sortieren läßt, produziert er wieder genau dieselbe Aufteilung. Fred sagt uns, daß für ihn Rot_1 und Rot_2 genauso unterschiedlich sind wie, sagen wir, Gelb und Blau. Er hat oft versucht, seinen Freunden den Unterschied zwischen Rot_1 und Rot_2 zu erklären, ist aber mit der Zeit zu dem Eindruck gelangt, daß alle anderen Menschen Rot_1-Rot_2-blind sind. Tatsächlich stellen wir fest, daß Freds Gehirn in einer differenzierteren Weise auf elektromagnetische Schwingungen in einem bestimmten Teil des elektromagnetischen Spektrums reagiert und so sein behaviorales Profil bezüglich bestimmter Diskriminationsleistungen erweitert. Aber wieviel wir auch immer über sein Gehirn und die ihm möglichen Zustände wissen werden, wir werden nie verstehen, wie die neue Farbe aussieht. Unser physikalisches Wissen über seine Erlebnisse bleibt, so scheint es, aus prinzipiellen Gründen bruckstückhaft, weil es das fragliche *Quale* ausklammert.

Wesentlich bekannter als Fred ist Mary (zumindest in Kreisen von Philosophen des Geistes). Mary ist eine brillante Neurowissenschaftlerin, die gezwungen war, ihr ganzes Leben in einem schwarz-weißen Raum zu verbringen und ihr wissenschaftliches Studium über einen sie mit der Außenwelt verbindenden Schwarzweißmonitor[23] zu absolvieren. Mary hat sich

---

anderen Perspektive attackieren Nemirow 1980, 1990 (bezüglich der Beziehung von Qualia und funktionalistischen Analysen siehe auch Nemirow 1979) und Lewis 1983 die Äquivokation von „Wissen" im Jacksonschen Argument.

[22] Ich habe bereits darzulegen versucht, was für eine Art von Wissen dies sein könnte. Jackson selbst beschreibt die epistemischen Intuitionen von „Qualia Freaks" folgendermassen: *„I think that there are certain features of bodily sensations especially, but also of certain perceptual experiences, which no amount of purely physical information includes. Tell me everything physical there is to tell about what is going on in a living brain, the kind of states, their functional role, their relation to what goes on at other times and in other brains, and so on and so forth, and be I as clever as can be in fitting it all together, you won't have told me about the hurtfulness of pains, the itchiness of itches, pangs of jealousy, or about the characteristic experience of tasting a lemon, smelling a rose, hearing a loud noise or seeing the sky."* Vgl. Jackson 1982: 127.

[23] Mit einem *„Butterfield encoder"*, also einem auf die richtige Weise animierten Schwarzweißbild, könnte man Mary allerdings jederzeit subjektive Farberlebnisse vermitteln, weil allein die Präsentation von Schwarzweiß-Signalen in der richtigen Pulssequenz zur Erzeugung *subjektiver* chromatischer Effekte ausreicht. Vgl. Butterfield 1968, 1970; Hardin 1988: 72.

alle physische Information[24] über den visuellen Cortex, die Stimulation der Netzhaut durch verschiedene Kombinationen von Wellenlängen, die kortikale Entstehung von sich auf Farberlebnisse beziehenden Sprechakten usw. angeeignet. Was geschieht nun, wenn Mary erstmals ihr achromatisches Gefängnis verläßt und eine rote Tomate oder den blauen Himmel sieht? Erfährt sie etwas *Neues* über die Welt und unsere visuelle Wahrnehmung derselben?

Die Konklusion des *Knowledge-Arguments* beruht darauf, daß Mary in der Tat etwas über die Welt und die inneren Zustände anderer Menschen erfährt, das sie vorher nicht wissen konnte. *Ex hypothesi* besaß Mary jedoch ein komplettes physikalisches Wissen über ihre Mitmenschen und deren Farberlebnisse. Darum muß der Physikalismus falsch sein: Es gibt Eigenschaften mentaler Zustände, denen gegenüber eine physikalische Beschreibung der Welt indifferent bleiben muß. Daraus folgt für Jackson ebenfalls, daß Qualia bestenfalls Nebenprodukte der Evolution unserer Nervensysteme darstellen, welche keinerlei abwärtsgerichtete kausale Wirksamkeit besitzen und irrelevant für wissenschaftliche Verhaltenserklärungen sind. Und deshalb wäre es auch falsch, zu meinen, es wäre die *Schmerzhaftigkeit* von Schmerzen, die uns etwa unsere Hand von einer heißen Herdplatte zurückziehen läßt – zu dieser Annahme kommen wir nur, weil wir biologische Systeme mit einem relativ niedrigen Grad interner Selbstrepräsentation sind. Jackson schreibt:

> We are the products of Evolution. We understand and sense what we need to understand and sense in order to survive. Epiphenomenal Qualia are totally irrelevant to survival. At no stage of our evolution did natural selection favor those who could make sense of how they are caused and the laws governing them, or in fact why they exist at all. And that is why we can't.[25]

Das *Knowledge-Argument* krankt natürlich an einer klaren Äquivokation von „Wissen" (und ist auch unter diesem Gesichtspunkt kritisiert worden), weil propositionales Wissen durch die Sätze wissenschaftlicher Theorien etwas vollkommen anderes ist als nicht-diskursives Wissen durch spezielle Formen interner Darstellung in menschlichen Gehirnen. Es lenkt aber unsere Aufmerksamkeit auf die Frage der Epistemizität subjektiver Zustände und ermöglicht uns, genauer zu diskutieren, *welche* Art von Information

---

[24] Mary besitzt eine enge Verwandte, nämlich Jeff Foss' *Super Neuro-Scientist* (SNS). In einem gegen Nagel gerichteten Gedankenexperiment versucht Foss zu demonstrieren, daß sowohl die SNS als auch ihr fiktiver Rivale, der SS (= *Super Sympathist*, ein Wesen mit maximalen empathischen Fähigkeiten, daß dank seiner psychischen Plastizität jedes andere bewußte Wesen phänomenal emulieren kann) die inneren Erfahrungen des Autors genausogut zu beschreiben vermag, wie er selbst – durch die Voraussage von Sprechakten oder durch qualitatives Einfühlen (vgl. Foss 1987). Foss schreibt: *„It cannot be shown (or at least remains to be shown) that there is anything of epistemic interest, whether information, ability, or a special species of knowledge, which SS has and SNS has not, and which is either implied by the hypothesised differences between them, or analytically contained in the hypothesis."* (Foss 1987: 217)

[25] Vgl. Jackson 1982: 134.

ein System über den qualitativen Gehalt mentaler Zustände erwirbt. Das Wissen, das Mary durch ihre neuen Erlebnisse über die Erlebnisse *anderer* Menschen gewinnt, ist ein Wissen jenseits von Wahrheit und Falschheit. Eine naturalistische Theorie des Subjekts muß uns deshalb die Beziehung zwischen qualitativem „Knowledge by Acquaintance" und wahrheitsfunktionalem „Knowledge by Description" bezüglich eben dieser phänomenalen Zustände erklären.

Kehren wir nach diesem kurzen Blick auf das Problem privaten Wissens zurück zur Relationalität des psychischen Phänomens Subjektivität. Subjektivität bezeichnet eine *Beziehung*: Mentale Zustände werden zu subjektiven mentalen Zuständen, indem sie in diese Beziehung eintreten. Sie treten in eine Beziehung ein zu einer Entität, die begriffsgeschichtlich aus idealistischen Konzeptionen des Selbstbewußtseins stammt und deren psychische Funktion alles andere als gut verstanden ist – besonders was ihre Beziehung zu theoretischen Entitäten in empirischen Theorien über das Gehirn und von ihm erbrachte kognitive Leistungen angeht. Diese Entität wird üblicherweise als das Subjekt bezeichnet.

Wenn Thomas Nagel uns nun sagt, daß er wissen möchte, wie es für eine *Fledermaus* ist, eine Fledermaus zu sein, dann vermischt er dabei die Frage nach der Identität des Subjekts mit der des qualitativen Gehalts seiner Zustände. Häufig oszillieren seine Bemerkungen zwischen diesen beiden Fragestellungen. Was aber ist es für eine Beziehung, durch die das psychologische Subjekt seine Identität fixiert, indem es manche mentalen Zustände zu *seinen* macht? Wir wissen aus empirischen Untersuchungen, daß die fragliche Subjektbeziehung mentaler Zustände durch rein naturwissenschaftlich zu beschreibende Ereignisse in ihrer Existenz determiniert ist: Sie entsteht – beim Aufwachen und in Fällen gelungener Psychotherapie –, sie kann verlorengehen – bei geistigen Erkrankungen aller Art, unter Narkose oder unter Einwirkung psychoaktiver Substanzen – und sie kann sogar multipliziert werden, nämlich in den bereits erwähnten Fällen von multipler Persönlichkeit (**MPD**). Daraus folgt unmittelbar, daß es auch nichtsubjektive mentale Zustände gibt. Wenn wir also verstehen möchten, wie die phänomenale Gesamtqualität des „Wie es ist, ein X zu sein" in die Welt kommen kann und konnte, müssen wir uns zusätzlich dem Ich oder Selbst zuwenden und es in seiner Funktion als Subjekt – das heißt: in seiner Relationalität – analysieren. In neueren Veröffentlichungen ist auch Thomas Nagel diesen Weg gegangen.

4.2.2 Perspektivische Fakten

Die vorläufig letzte Variante der Problemexposition stellt Thomas Nagel in seinem Buch *„The View from Nowhere"* vor. Obwohl auch hier die Kernthese eine schillernde ist, läßt sie sich doch kurz folgendermaßen charakterisieren: *„Es gibt Wahrheiten, die nur in der ersten Person Singular ausgesprochen werden können. Solche Wahrheiten beziehen sich auf eine be-*

stimmte Klasse von Tatsachen – perspektivische Fakten – ,die im wissenschaftlichen Weltbild notwendigerweise fehlen müssen.“ Werfen wir einen Blick auf die Art und Weise, in der Nagel das Problem entwickelt, um zu sehen, welche neuen Aspekte er der Frage nach dem Wesen des Subjekts abgewinnt.

Überraschend ist an Nagels Reformulierung des Problems, daß es ihm nicht mehr um die Qualität des „Wie es ist, ein X zu sein“ geht, sondern um eine bestimmte *Tatsache*. Dieser in der Analyse von Subjektivität als einem Faktum bestehende philosophische Schachzug Nagels ist von einem seiner wichtigsten Kritiker polemisch als *move to funny facts* bezeichnet worden.[26] Was bedeutet es, *jemand zu sein*? Jeder einzelne von uns wäre – selbst wenn alle Erlebnisperspektiven und subjektiven Standpunkte Teile einer Beschreibung der Welt geworden sind – mit dem Problem partikularer Subjektivität konfrontiert: In einer Totalrepräsentation der Welt käme – so sagt Nagel – eine wichtige Tatsache niemals vor, und zwar die Tatsache, daß *ich* mit einer der historischen Personen in ihr identisch bin. Es gibt nämlich eine große Zahl von Objekten und Erlebnisperspektiven in dieser Welt, einer Welt, die ihrer objektiven Beschreibung nach mittelpunktlos ist. Objektiv beschriebene Welten können nur räumliche Mittelpunkte haben, keine phänomenalen. Das subjektive Bild der Welt dagegen scheint (obwohl es schwerlich als „Beschreibung“ im engeren Sinne gelten kann) essentiell zentriert zu sein und überhaupt keinen Ort im Raum der Physik zu besitzen – wenn man nicht annimmt, daß es in einem strikten Sinn mit gewissen neuronalen Zuständen identisch ist. Aber selbst wenn der fragliche intentionale und qualitative Gehalt des subjektiven Bilds der Welt auf neurowissenschaftlich zu beschreibende Zustände reduziert werden könnte, scheint seine Zentriertheit eine Eigenschaft zu sein, die wir schwerlich auf der physischen Seite der postulierten Identitätsrelation wiederfinden werden. Eine solche vollständige Darstellung der Welt läßt darum die für jeden von uns zugleich wichtigste und rätselhafteste Frage offen: Wie kann es möglich sein, daß *ich* eine dieser Einzelpersonen (inklusive aller ihrer öffentlichen Eigenschaften und ihres subjektiven Standpunkts) bin? Wie kann *ich* TM sein, in einer Welt, deren Brennpunkt ich bin, die jedoch andererseits als absolut zentrumlos erscheint? Wie ist es möglich, daß ich *qua zentriertes Bewußtsein*[27] mit einem einzelnen Objekt, mit einer partikularen Person in einer mittelpunktlosen Welt identisch bin?

Es scheint, als ob der subjektive und der objektive Standpunkt auf ewig unversöhnbare Gegensätze bilden müßten. Nun unterscheidet Nagel zwei mögliche Fragen hinsichtlich der intuitiven Inkompatibilität der beiden Weisen des Wirklichkeitszugangs. **Erstens**: Wie kann es von einer bestimmten Einzelperson, die nur eine von vielen in einer zentrumlosen Welt ist, wahr sein, daß sie *ich* ist? *Prima facie* scheint dies eine wichtige Wahrheit

[26] Vgl. Lycan 1987: 77.
[27] Vgl. Metzinger 1989.

betreffs des fraglichen Individuums zu sein – und diese Wahrheit würde in einem wissenschaftlichen Weltbild fehlen und es so prinzipiell unvollständig machen. **Zweitens** – vom Subjekt aus gefragt –: Wie ist es möglich, daß ich nichts als diese individuelle Person mit ihren begrenzten Eigenschaften und ihrer zufälligen Biographie bin? Es scheint etwas in mir zu geben, das es mir unmöglich macht, die Aussage „Ich bin TM" als eine Identitätsaussage zu lesen und zu verstehen. Mit anderen Worten: Es gibt ein essentielles Moment in meinem Innenleben, das es mir nahelegt „Ich bin TM" als eine Subjekt-Prädikat-Proposition zu interpretieren. Dieses fragliche Moment macht es mir unmöglich, zu glauben, daß ich als denkendes und fühlendes Subjekt aufgehen soll in einem Bündel zufälliger Eigenschaften.

Mit diesem zweiten Frageaspekt beutet Thomas Nagel eine weitere Intuition aus, die bei den meisten Menschen des westlichen Kulturkreises stark ausgeprägt ist und die ich ab jetzt die „Kontingenz-Intuition" nennen werde: *„Ich hätte auch ein(e) ganz andere(r) sein können."* Viele von uns meinen, daß sie als Personen auch ganz andere öffentliche Eigenschaften hätten besitzen können und daß ein Teil von ihnen sich dabei nicht verändert hätte. Die Kontingenz-Intuition verleiht uns das Gefühl, als gäbe es einen unveränderlichen Wesenskern in uns, dessen öffentliche Eigenschaften einem ständigen und mehr oder weniger akzidentiellen Wandel unterworfen sind.[28] Beim Übergang von einem objektiven in ein subjektives Weltbild erscheint uns die Identifikation mit einer bestimmten historischen Person dann als kontingentes Ereignis: Wir sind nicht notwendigerweise zu diesem oder jenem Individuum geworden. Und deswegen scheint uns, als sei eine wichtige und essentielle Wahrheit bezüglich unserer selbst mit einer solchen quasi-empirischen Identitätsaussage niemals auch nur tangiert.

Eine perspektivelose und objektive Repräsentation der Welt, der sich in ihr befindenden Personen und ihres Mentallebens – wie sie vielleicht eine an ihr historisches Ende gekommene Wissenschaft darstellen könnte – läßt immer eine wichtige Wahrheit über *eine* der unzähligen Personen aus, sagt Nagel. Es ist die Wahrheit, daß diese Person der locus *meines* Bewußtseins und der Mittelpunkt *meiner* Welt ist. Diese Wahrheit, ausgedrückt in der ersten Person Singular, spiegelt eine Tatsache wider. Es gibt irreduzible Erste-Person-Fakten, sie können nur durch den selbst-lokalisierenden Gedanken oder Satz „Ich bin TM" ausgedrückt werden und es hat den Anschein, als ob diese Fakten sich nicht in der Welt befänden.

---

[28] Diese Intuition liegt natürlich auch vielen zentralen Denkfiguren der klassischen, spekulativ-idealistischen Subjektphilosophie zugrunde. Manfred Frank schreibt: *„Nagel bringt einige Grundintuitionen der frühidealistischen deutschen Philosophie in Erinnerung und in Geltung, z. B. Fichtes (etwas umformuliertes) Problem, daß es keine von zeichenreflexiven Ausdrücken (wie ›hier‹, ›jetzt‹, ›ich‹) freie Beschreibung gibt, aus der erhellt, daß ich diese Person bin. Das führt zu einer Rehabilitierung des Gewißheitselements, in dem mir und nur mir meine mentalen Erlebnisse zugänglich sind, und zur Feststellung einer unüberwindlichen ›epistemischen Asymmetrie‹ zwischen Beschreibungen meiner durch mich selbst und solchen meiner durch andere."* (Frank 1991: 177)

Unter vielem anderen mag man an dieser Stelle einwenden, daß es sich hier um eine sprachliche Verhexung handelt. Nagel jedoch geht es nicht darum, ein Scheinproblem zu konstruieren und durch essentialistische Intuitionspumpen[29] zu plausibilisieren – er mutet seinen Lesern zu, unter der sprachlichen Oberfläche ein reales Problem zu erkennen.

> Solche Fragen mögen dem Leser selbst dann lächerlich vorkommen, wenn er sie sich in bezug auf sich selbst stellt; ich versuche hier jedoch, ein geschärftes intuitives Problembewußtsein zu wecken und ihn davon zu überzeugen, daß auch dann etwas daran ist, wenn die sprachliche Darstellungsform unvollkommen sein mag. Es mag Fälle geben, in welchen eine sprachliche Taschenspielerei uns das Trugbild einer Frage vorgaukelt, wo in Wahrheit keine Frage besteht, aber dies ist kein solcher Fall. Wir können die Frage auch unabhängig von ihrem verbalen Ausdruck in unserem Denken wahrnehmen, und die Schwierigkeit besteht gerade darin, sie so zu stellen, daß sie nicht zu einer Oberflächlichkeit wird oder dazu einlädt, daß man sie nach Maßgabe ihrer verbalen Form beantwortet, ohne wirklich dem Problem unter der Oberfläche gerecht zu werden. In der Philosophie handelt es sich niemals bloß darum, wie wir uns ausdrücken sollen.[30]

Ich glaube, man kann Nagel konzedieren, daß er ein bedeutsames und nicht-triviales Problem umkreist. Vielleicht kann man die von ihm aufgeworfenen Fragen als Ausdruck eines durch philosophische Introspektion ausgelösten Staunens über die Komplexität unserer eigenen inneren Natur auffassen. Die Selbstidentifikation mit einer historischen Person durch einen Sprechakt ist in der Tat ein erstaunlicher Vorgang – vor allem deshalb, so meine ich, weil er mit einer *mentalen* Operation verknüpft ist und mit einer phänomenalen *Zustandsänderung* einhergeht. Dieser Vorgang wird um so erstaunlicher und geheimnisvoller, wenn wir ihn introspektiv zu beobachten versuchen. Was geschieht eigentlich in dem Moment, in dem ich – sei es nun extern oder mental – den Satz „Ich bin TM“ äußere? Introspektiv beobachtet, stellt die affirmative Äußerung des Satzes „Ich bin TM“ ein mentales Ereignis dar: Etwas verändert sich in mir, sobald ich diesen Satz ausspreche. Bevor ich den kritischen Satz – den Nagel auch als „philosophisch“ und „selbst-lokalisierend“ bezeichnet – denke oder spreche, repräsentiere ich die Welt von einem objektiven Standpunkt aus, einschließlich aller psychologischen Eigenschaften und Perspektiven aller mir bekannten Personen im Universum. Alle denkbaren subjektiven Beschreibungen meiner selbst und anderer Personen sind verobjektiviert worden, gewissermaßen aus dem Ich hinausgeschoben in eine Totalansicht der Welt *sub specie aeternitatis*. In dieser Totalansicht der Welt ist Thomas Metzinger (höchstens) das, was Nagel treffend als „eine Person unter zahllosen anderen in Ozeanen von Raum und Zeit“[31] oder als „momentanes Echozeichen auf dem kosmischen Bildschirm“[32] bezeichnet. Diese Totalansicht

[29] Dieser Ausdruck stammt von Daniel Dennett. Vgl. etwa Dennett 1988.
[30] Vgl. Nagel 1992 (1986): 99.
[31] Vgl. Nagel 1986: 61.
[32] Vgl. Nagel 1986: 61

wird durch den *Blick von nirgendwo* erzeugt, sie bringt eine vollkommene Relativierung von TM und eine Loslösung von seinem Standpunkt in der Welt mit sich. In dem Moment, in dem ich mich dann durch Selbstidentifikation in meinem inneren Bild der Wirklichkeit selbst lokalisiere, verändert sich dieses Bild – eine phänomenale Zustandsveränderung tritt ein. Durch das Aussprechen des Satzes „Ich bin TM" verändere ich somit psychologische Eigenschaften meiner selbst und transformiere gleichzeitig ein aperspektivisches inneres Bild der Welt in ein subjektzentriertes. Dieser Übergang ist dramatisch, das Staunen und die philosophische Analyse sind seiner mit Sicherheit wert. Ich werde darum im Schlußkapitel eine Antwort auf die Frage geben, was geschieht, wenn wir aus dem unendlichen Raum, den der *View from Nowhere* in uns öffnet, abstürzen in die Begrenzungen und Perspektivengebundenheit eines partikularen psychologischen Subjekts.

An dieser Stelle scheint mir zunächst der Hinweis wichtig, daß das von Nagel vorsichtig entwickelte Problem nicht verkürzt werden sollte zu einem bloß *semantischen* Problem, zu einer Frage von Selbstreferenz oder der sprachlichen Logik von Eigennamen. Die analytische Behandlung unseres sprachlichen Zugriffs auf die fraglichen mentalen Phänomene darf die Einsicht nicht unterschlagen, daß eine *epistemische* Asymmetrie verborgen ist hinter dem, was Manfred Frank als das „Prinzip der semantischen Symmetrie" zwischen Selbst- und Fremdzuschreibungen von Bewußtseinsprädikaten bezeichnet hat.[33] Außerdem müssen wir versuchen, uns auf dem Weg über eine Naturalisierung den Phänomenen wieder anzunähern: Wir haben es bei dem Problem der Subjektivität mentaler Zustände mit einem interessanten und schlecht verstandenen psychischen Phänomen zu tun. Anscheinend gibt es unterschiedliche Weisen, in denen wir uns selbst intern repräsentieren können – zumindest auf der Ebene intendierter Imagination.[34] Intendiertes Imaginieren ist – genau wie Introspektion – eines der

[33] Vgl. etwa Frank 1991: 33. Das, was uns an phänomenalem Selbstbewußtsein eigentlich interessiert, ist das *präreflexive* Element: Es kann nicht durch einen Sprachapriorismus oder durch aus dem intersubjektiven Raum heraus operierende Erklärungsversuche aufgefangen werden. In Franks Worten: „ *. . .wäre Selbstbewußtsein an die Sprachkompetenz gebunden (so, daß ich ›ich‹ nur erlernen könnte, wenn ich zuvor die Möglichkeit der Identifizierbarkeit aus der ›er‹-Perspektive begriffen und die Konvertierbarkeit der Perspektiven verinnerlicht hätte), so würde meine unbedingte (cartesianische) Selbstvertrautheit gerade an eine Bedingung gebunden: sie hinge ab von der Verinnerlichung von Regeln, die mich als Inter-Subjekt konstituierten, noch bevor ich Gelegenheit hatte, ein Subjekt zu sein. Die ganze Weisheit tradierter Einsichten in scheiternde Explikationsversuche von Selbstbewußtsein resümiert sich aber in der schlichten Sentenz, daß von Fremdem (und von Fremden) nie zu lernen ist, daß* ***ich*** *dieses (oder dieser) Fremde bin (wenn ich es nicht schon vorher wußte). Die Vorschaltung der ›er‹-Perspektive vor der ›ich‹-Perspektive fügt den vielen Zirkeln in der Selbstbewußtseins-Erklärung nur eine neue Variante hinzu.*" (Frank 1991: 36)

[34] Intendiertes Imaginieren ist ein Fall von mentaler *Simulation* (vgl. Abschnitt 2.1.2). Nagel unterscheidet häufig nicht zwischen dem *View from Nowhere* als einer absichtlich herbeigeführten inneren Simulation und als *repräsentationalem Gesamtzustand* (vgl. Abschnitt 2.3.1), d. h. als *Bewußtseinszustand.* Letzterer wäre eine erlebnismäßig unhintergehbare „mystische Erfahrung", ersteres jedoch ist nicht mehr als ein philosophisches Gedankenexperiment. Der *Blick von nirgendwo* ist – eine wichtige Differenzierung – also eine *Operation*

wichtigsten philosophischen Instrumente und eines der gefährlichsten dazu, weil die Fähigkeit zur Imagination eine natürliche Fähigkeit mancher Biosysteme ist, die unter hochspezifischen Bedingungen zu hochspezifischen Zwecken entwickelt worden ist. Wie bereits gesagt ist Imaginierbarkeit durch menschliche Gehirne aber nicht dasselbe wie Analytizität, weil Imagination – als kognitive Leistung betrachtet – wahrscheinlich eine große Anzahl „impliziter Annahmen" des Gehirns über das Wesen der Wirklichkeit reproduziert, die dem bewußten Erleben nicht zugänglich sind. Auf dem Gebiet der Philosophie des Geistes ist die Strategie des imaginativen Gedankenexperiments besonders problematisch, weil sich Imagination und Introspektion zu leicht nach dem Modell sinnlicher Wahrnehmung miteinander vermischen.[35]

Es ist aber der Einsatz dieses Instruments, den Nagel seinen Lesern zumutet, wenn er sagt, daß wir alle das philosophische Problem unabhängig von seiner verbalen Formulierung auch *fühlen* können. Es mag jedoch sein, daß wir unterschiedlich fühlen oder daß einige von uns nur *ein* Problem fühlen, wo andere *mehrere* von ihnen spüren. Wir müssen deshalb fragen: Was ist es genau, das wir fühlen, wenn wir Sätze vom Typ „Ich bin TM" aussprechen? Welche mentalen Ereignisse gehen mit dieser Äußerung einher und was verändern sie an uns selbst? Worüber berichten wir, über die selbstreferentielle Bezugnahme auf einen durch eine Reihe publiker Eigenschaften charakterisierten Sprecher hinaus? Ist es denkbar, daß wir mit Sätzen des problematischen Typs faktisch in unreflektierter und verworrener Weise versuchen, auf Änderungen unseres Bewußtseinszustandes bezugzunehmen und sie mitzuteilen ? Ich werde diese Fragen noch zu beantworten versuchen. An diesem Punkt unserer Überlegungen geht es mir nur um den Hinweis, daß wir nicht bloß einem sprachanalytisch interessanten Problem von Subjektivität, Selbstreferenz und angeblich irreduziblen Erste-Person-Fakten gegenüberstehen, sondern auch einer Frage bezüglich mentaler Repräsentation und vor allem mentaler Selbstrepräsentation.

Wenn unsere Gehirne nicht in der Lage wären, ganz bestimmte Formen interner Selbstrepräsentation zu erzeugen, dann verlöre auch Nagels Problemexposition sofort ihre intuitive Überzeugungskraft. Wir wären dann nicht mehr in der Lage, das Problem jenseits aller sprachlichen Formulierungen zu fühlen. Vielleicht kann man das präreflexive, empathische Erfühlen und Aufspüren von philosophischen Problemen selbst zu einem Objekt von Kognitionsforschung machen und so herausfinden, warum manche philosophischen Fragestellungen (und viele schlechte Antworten) notorisch populär sind. Das könnte uns möglicherweise auch *vis à vis* des hier zentralen Fragenkomplexes weiterhelfen. Denn was Nagels Problemformulierung so einleuchtend macht, sind mit Sicherheit nicht nur ins

---

und sicherlich eine unserer interessantesten mentalen Kapazitäten, aber keine Einswerdung mit einem transzendentalen Ego. Vgl. auch Abschnitt 5.2.

[35] Gegen den Einsatz von Gedankexperimenten in der Philosophie des Geistes argumentiert Kathy Wilkes in Wilkes 1988. Vgl. dazu auch Buschlinger 1993.

Alltagsidiom abgesunkene cartesianische Intuitionen, sondern auch spezifische Eigenschaften der Informationsverarbeitung in unseren Gehirnen. Das ist so, weil Intuitionen nicht nur im Zusammenhang öffentlicher Sprachpraxis und geistesgeschichtlicher Traditionen entstehen, sondern auch vor dem Hintergrund der sehr speziellen evolutionsbiologischen Entstehungsbedingungen des Informationsflusses im Gehirn.

Festzuhalten bleibt, daß der *Blick von nirgendwo* eine notwendige Vorbedingung für das Auftreten einer bestimmten Klasse von mentalen Ereignissen ist. Ohne diese spezifischen mentalen Ereignisse ließe sich Nagels Problem der psychologischen Subjektivität nicht formulieren. Die mittelpunktlose Abbildung der Welt ist die *conditio sine qua non* für das Entstehen des philosophischen Problems und Nagels Behandlung seines **zweiten** Aspekts – *Wie kann ich nichts weiter als eine partikulare Person sein?* – läßt die Bedeutung der nicht-zentrierten internen Weltrepräsentation noch deutlicher werden.

> Es ergibt sich das folgende Bild: Meinem Wesen nach habe ich schlechterdings gar keine besondere Perspektive, sondern fasse die Welt azentrisch auf. Kontingenterweise schaue ich die Welt gewöhnlich aus einem bestimmten Blickwinkel an und benutze dabei die Augen, die Person, das Alltagsleben von TN als eine Art Fenster. Doch die Erlebnisse und die Perspektive von TN, die mir unmittelbar gegeben sind, machen nicht den Standpunkt des eigentlichen Selbst aus, da das eigentliche Selbst eben keine Perspektive hat, sondern TN und seine Perspektive in seiner Auffassung von einer zentrumlosen Welt als Inhalte dieser Welt miteinschließt. Dieser Aspekt des Selbst ist es, der gemeint ist, wenn ich die Welt als eine Gesamtheit denke und frage: „Wie kann TN ich sein? Wie kann ich TN sein?"; und dieser Aspekt ist es, dem der sich selbst lokalisierende philosophische Gedanke seinen eigentümlichen Gehalt verdankt.[36]

Das hier eingeführte objektive Selbst trägt starke metaphysische Züge: Es wird uns als das *wahre* Selbst vorgestellt und ist Gegenstand einer Wesensaussage, welche es als essentiell perspektivelos präsentiert. Eine naturalistische Theorie des Geistes wird also – vorausgesetzt sie akzeptiert die zugrundegelegte phänomenologische Analyse der inneren Situation – erklären müssen, *warum* uns das objektive Selbst als das wahre Selbst erscheint und warum wir den introspektiven Eindruck gewinnen, dieses Selbst könnte niemals ein anderes gewesen sein als ein perspektiveloses Selbst und sei darum logisch zu unterscheiden von allen anderen Formen erlebter psychischer Identität. Nagel gibt zu, daß die alleinige Existenz derjenigen mentalen Instanz, die er als das objektive Selbst bezeichnet, noch nicht die Zufälligkeit der Verbindung zwischen ihm qua objektives Selbst und der historischen Person TN beweist. Es zeigt jedoch – so Nagel – daß etwas Essentielles an uns nichts mit unserer Perspektive und Position in der Welt zu tun hat.[37]

---

[36] Vgl. Nagel 1992 (1986): 109.
[37] Vgl. Nagel 1992 (1986): 109f.

Dieses scheinbar essentielle Merkmal von menschlichen Subjekten nun ist Nagels eigentliches Thema. Vielleicht kann man das von ihm angedeutete Projekt als Rekonstruktion des Husserlschen transzendentalen Egos ohne den Hintergrund eines transzendentalen Idealismus beschreiben.[38] Das von ihm postulierte objektive Selbst ist das letzte Stadium des sich von der Welt und seinen eigenen Bewußtseinsinhalten ablösenden Subjekts, bevor es zu einem ausdehnungslosen Punkt zusammenschrumpft. Eventuell könnte man es auch als eine auf der Grenze zwischen Naturalismus und Essentialismus angesiedelte theoretische Entität beschreiben: Sie soll eine unserer dunkelsten und faszinierendsten psychologischen Eigenschaften erklären und ist dabei ständig in Gefahr, entweder in eine außerweltliche, transzendentale Leere abzugleiten oder sich in Form der Erlebnisperspektive einer besonderen, innerweltlichen Person zusammenzuziehen.

Wie konstruiere ich aus der Person TM ein objektives Selbst? Ich tue es, indem ich alle meine psychischen Zustände – ebenso wie meine öffentlichen Eigenschaften – als Rohmaterial für die Erzeugung einer objektiven Projektion meiner selbst einsetze. Ich behandele Informationen, die mir normalerweise im Modus direkter Gegebenheit zur Verfügung zu stehen scheinen, wie indirekte Quellen *externen* Wissens. Ich bereichere die objektive Welt um jene Eigenschaften und Zustände, die ich bis jetzt immer als meine Zustände bezeichnet habe, weil ich sie ab jetzt nur noch *von außen* betrachte. Indem ich dies tue, scheint etwas zu entstehen, das Philosophen traditionell fasziniert hat: ein reines, inhaltsloses Ich. Denn eine anspruchsvolle Theorie der Subjektivität muß uns erklären können, ob es andere Formen der Selbstidentifikation gibt als durch öffentliche Eigenschaften, als durch sprachliche Bezugnahme auf die eigene Biographie. Kann es *reine* – d. h. gehaltfreie – Formen mentaler Selbstreferenz geben oder ist das *inhaltslose Ich* in Wirklichkeit nur eine Form von *ichlosem Inhalt*?

---

[38] Die zweite verwandte Position, auf die Nagel sich explizit bezieht, ist Wittgensteins metaphysisches Subjekt (TLP 5.641). Auch dieses Subjekt ist die logische Grenze einer Welt, Nagel möchte jedoch den mit ihm verknüpften Solipsismus und seine Transzendentalität eliminieren. Von Michael Gebauer habe ich gelernt, daß Nagel, insofern man seine Position als einen Typ von neutral-monistischer Doppelaspekt-Theorie interpretieren kann, viel dem Denken Baruch de Spinozas verdankt. In den profunden unveröffentlichten Nachbemerkungen zu seiner deutschen Übersetzung von Nagels *View from Nowhere* schreibt Gebauer bezüglich der von mir im Text angedeuteten Verwandtschaft des Husserlschen mit dem Nagelschen Begriff des Selbstbewußtseins: *„Bemerkenswert ist auch die buchstäbliche Affinität des § 11 („Das psychologische und das transzendentale Ich") von Husserls Cartesianischen Meditationen zu dem hier bei Nagel in allen Kapiteln präsenten Gedanken von der Person als einem „Stück der Welt" in Wittgensteins Sinne, . . ., zu dem vom Autor . . . in Anspruch genommenen Gedanken der „Teilung des Selbst" in ein „interessiertes Ich" und einen „anteilnahmslosen Zuschauer". . . . scheint Nagel allerdings darauf hinzuweisen, daß es gerade keiner* ***transzendental****-phänomenologischen Reflexion bedarf, um den Aspekt des Subjekts, den er hier im Gegensatz zur empirischen Person das „objektive Selbst" nennt, von dieser Person zu abstrahieren, und dieser Hinweis ist wichtig, wenn Nagel auch im Kontext seines in einem ganz anderen Sinne nichtempirischen Subjekts jede Variante des Idealismus, auch die Wittgensteinsche einer singulären und überpersönlichen objektiven Weltseele als mit dem Bewußtsein seiner Individualität a priori unvereinbar ausschließen will."* (Gebauer 1992: 15)

Nagel sagt uns, daß sein objektives Selbst das letzte Stadium eines sich ablösenden Selbst ist, bevor dieses zu einem ausdehnungslosen Punkt zusammenfällt. Es gibt also einen imaginativen *Prozeß*, der von einem zentrierten Bewußtsein über ein objektives Selbst zu einem extensionslosen Punkt führt. Einerseits findet sich der fragliche psychische Prozeß hier nur in metaphorischer Beschreibung durch Anleihen in der Geometrie und bei visuell-räumlichen Modellen der Realität[39], andererseits eignet er sich gerade durch seine Prozessualität für eine naturalistische Analyse unter Einsatz unverdächtiger Kategorien wie der der mentalen Repräsentation. Die Entstehung dessen, was Nagel als das objektive Selbst in Anschlag bringt, ist ein reversibler mentaler Vorgang: eine prinzipiell beliebig oft wiederholbare psychische Operation. Wir wissen heute, daß psychische Prozesse aufs engste mit Informationsverarbeitungsereignissen im Gehirn zusammenhängen und dies eröffnet uns eine historisch neue Möglichkeit, Subjektivität naturalistisch zu analysieren. Unsere Fähigkeit, innere Perspektivenwechsel durchzuführen und dabei auch unseren Bewußtseinszustand zu verändern, ist offen für empirische Untersuchungen, sobald wir sie begrifflich als einen Fall von mentaler Repräsentation analysiert haben. Insofern kann die Aufgabe des Philosophen in bezug auf das Problem der Subjektivität mentaler Zustände weit über den metatheoretischen Kommentar zu den Fortschritten der relevanten Einzeldisziplinen hinausgehen.

Wenn wir Nagel folgen und annehmen, daß jeder von uns qua potentielles Subjekt einer zentrumlosen Konzeption der Welt ein objektives Selbst ist, dann wird auch verständlich, daß dieses Subjekt als inhaltslose Grenze der Welt in dem Augenblick verschwindet, in dem ich die Zauberformel „Ich bin TM“ ausspreche, weil es sich sofort mit Gehalt zu füllen beginnt. Denn mit der Zuschreibung von personalen Eigenschaften geht die Emergenz einer Perspektive einher. Ich kann jetzt wieder unter dem gewohnten Aspekt partikularer Individualität auf mich Bezug nehmen – indem ich einen Teil der Zustände in der Welt wiedererkenne als *meine* Zustände, die mir auf eine mehr oder weniger mysteriöse Weise nicht-derivativ gegeben sind.

Den „philosophischen Geschmack“ erhält – so Nagel – der Gedanke „Ich bin TM“ durch die Tatsache, daß er neben der Referenz auf eine historische Person gleichzeitig indexikalisch Bezug nehmen kann auf ein objektives Selbst und so die objektive und subjektive Konzeption der Welt zusammenführt. Nichts an der objektiven Konzeption der Welt deutet darauf hin, daß sie überhaupt ein Subjekt in der Welt besitzt. Da es jedoch ein Subjekt *dieser* Konzeption gibt, kann dieses Subjekt auch eine der relativen innerweltlichen Perspektiven als ihr *Fenster in die Welt* benutzen. Es ist dieser

[39] Wir stoßen hier wiederum auf ein bekanntes Problem: Metaphern, die die visuelle Modalität unseres phänomenalen Bewußtseins einsetzen, um metaphysische „Verdinglichungen“ psychischer Vorgänge intuitiv plausibel zu machen, sind deshalb so einleuchtend, weil unser physischer Sehapparat tatsächlich Mengen von Eigenschaften intern *als Objekte* darstellt. *„Das optische Modell des geistigen Auges reduziert ›sein‹ auf Gegenständlichsein.“* (Frank 1991: 188)

bemerkenswerte Zusammenhang, der in einer materialistisch-objektiven Kosmologie nicht untergehen darf. Ob eine naturalistische Anthropologie vor dem Hintergrund einer physikalistischen Kosmologie Nagels Thema in der hier vorliegenden Formulierung retten kann, scheint allerdings äußerst fraglich. Zu unklar ist der Status eines transzendentalen Ichs außerhalb einer idealistischen Rahmentheorie, zu sehr erinnert die historische Person als „Fenster zur Welt" an Entelechien oder die Drei-Welten-Lehre von Popper und Eccles[40], zu unklar ist eine Identitätsaussage auf deren einer Seite mit dem *objective self* eine eigenschaftslose Entität[41] steht. In einer solchen Situation kann man nur versuchen, über eine Kritik der fraglichen Position zu einem besseren Verständnis des philosophischen Problems und der psychischen Phänomene, in denen es gründet, zu gelangen. Man kann dann überprüfen, ob die so entstehende neue Theorie kompatibel mit einem wissenschaftlichen Weltbild ist oder ob auch in ihr das ursprüngliche philosophische Rätsel wieder auftaucht.

4.3 Schwierigkeiten der Nagelschen Subjektphilosophie

Ich werde nun einige mögliche Einwände gegen die Nagelschen Thesen zum phänomenalen Gehalt mentaler Zustände, zu ihrer Subjektzentriertheit und zur Existenz irreduzibler Erste-Person-Fakten betrachten. Dabei werde ich mich zuerst dem zuletzt diskutierten Aspekt zuwenden, der mit der Annahme eines objektiven Selbst zusammenhängt.

Nagel stellt eine Behauptung bezüglich einer bestimmten Klasse von Sätzen auf, Sätzen vom Typ „Ich bin TM". Diese Sätze sind ihrer logischen Form nach Identitätsaussagen und sie enthalten einen indexikalischen Ausdruck, nämlich das Wort „Ich". In ihrem Gehalt beziehen sich solche Sätze auf diejenige Person, die sie äußert: Es sind Fälle von *Selbstreferenz*. Nagel behauptet nun zweierlei: **Erstens** sind Sätze, in denen solche indexikalischen Ausdrücke vorkommen, nicht übersetzbar in objektive Aussagen, die von der dritten Person Gebrauch machen. **Zweitens** drücken diese Sätze eine nicht-triviale Wahrheit über uns alle aus, eine Wahrheit, die nicht in der Bezugnahme auf einen Sprecher mit einer bestimmten Geschichte und bestimmten öffentlichen Eigenschaften besteht.

---

[40] Vgl. Popper/Eccles 1982.

[41] Auch Norman Malcolm weist darauf hin, daß es keinerlei Identitätskriterien für das objektive Selbst gibt. Er schreibt: *„But in regard to the identity of an I that supposedly occupies the point of view of a person, we could be neither right nor wrong. After a bout of severe amnesia Nagel might be able to identify himself as TN – but not as I. ‚I am TN' could announce a discovery – but not ‚I am I'."* (Malcolm 1988: 160) Malcolm weist auch auf die frappierende Ähnlichkeit zwischen dem cartesianischen *moi* und dem Nagelschen *objective self* hin – wobei das letztere, wenn man Nagel ernst nimmt, auch keine *psychologischen* Eigenschaften besitzt. Vgl. Malcolm 1988: 154 und 159; dazu auch Metzinger 1995 d.

Wenn das wahr ist, dann muß es einen Typ von Faktum in der Welt geben, der das Referenzobjekt für den zweiten inhaltlichen Aspekt der fraglichen Identitätsaussagen darstellt (sie besäßen demnach einen „doppelten Set von Wahrheitsbedingungen"). Dieser Typ von Tatsache könnte sich nicht in aus objektiven Aussagen bestehenden Beschreibungen der Welt – wie sie die Wissenschaft liefert – wiederfinden. Nagel behauptet also für die fraglichen Sätze eine doppelte Referenz. Er tut dies aber, ohne die logische Struktur der von ihm postulierten Tatsachen anzugeben. Man könnte jetzt die doppelte Referenz auch als eine doppelte Relation des Subjekts zur Welt zu verstehen versuchen, der erste Ansatzpunkt einer Kritik wird trotzdem ein *semantischer* sein müssen.

Die offensichtlichen Wahrheitsbedingungen des Satzes „Ich bin TM" bestehen in einer Anzahl von Wahrheiten über die historische Person des Sprechers und in einem simplen Kontext der Äußerung: „Ich bin TM" ist nämlich genau dann wahr, wenn er von TM geäußert wird.[42] Er kann nicht substituiert werden durch eine Dritte-Person-Aussage, aber die seine Wahrheitsbedingungen ausmachenden Fakten können alle ausgedrückt werden durch solche Aussagen.

Indexikalische Ausdrücke wie „Hier", „Jetzt" oder „Ich" referieren auf einem bestimmten räumlichen, zeitlichen oder psychischen Standpunkt in der Welt, ohne daß der jeweilige Sprecher notwendigerweise auch ein *Wissen* über seine jeweilige Position besitzt: Man kann sich darüber irren, *wer* man ist, und selbstverständlich kann man sich auch über seine räumliche oder zeitliche Position im Universum täuschen. Die Tatsache, daß wir uns darüber täuschen können, wer wir sind, ist philosophisch nicht uninteressant. Denn ähnlich wie aus einer Totalbeschreibung des Universums durch eine an ihr Ende gekommene Physik nicht hervorginge, welcher Ort „hier" ist und welche Zeit „jetzt" ist, schiene aus einer vollständigen *psychologischen* Kosmologie für jeden einzelnen von uns nicht hervorzugehen, welches der vielen Systeme mit mentalen Zuständen *ich* bin.[43] Daß wir alle uns zum Beispiel während endogener Psychosen, unter Einfluß psychoaktiver Substanzen oder nach Amnesien darüber täuschen können, wer wir sind, zeigt, daß wir es in solchen Situationen mit einem gescheiterten Zugriff auf öffentliche Eigenschaften zu tun haben. Öffentliche Eigenschaften erzeugen

[42] Wenn TN unter dem Syndrom der multiplen Persönlichkeit (**DID**) leidet und eine seiner Subpersönlichkeiten den fraglichen Satz ausspricht, wodurch genau wird er dann falsch? Vgl. Abschnitt 3.2.2.

[43] In einer solchen „psychischen Landkarte" der Welt würde das fehlen, was viele zum Beispiel an den Wänden von U-Bahnhöfen fest installierte Stadtpläne von *tragbaren* (also in einer unendlichen Anzahl möglicher Situationen anwendbaren) Stadtplänen unterscheidet: ein kleiner roter Pfeil mit dem Hinweis „SIE SIND HIER" (vgl. Abschnitt 5.1). Die *Stärke* wissenschaftlicher Theorien besteht darin, daß sie universelle und nicht-zentrierte Repräsentationssysteme sind, die die Interessen individueller Benutzer außer acht lassen. *„The addition YOU ARE HERE to a map is, evidently, not the cartographer's business, and such maps could not be sold at bookstores; similarly the addendum THIS IS YOU to a completed psycho-physical account is not a proper part of the scientist's concern."* (Wilkes 1984: 241)

Fallibilität. *Private* Eigenschaften setzen dagegen – will man diesbezüglich überhaupt von der Möglichkeit privater Illusionen und Erkenntnisse sprechen – Descartes' starke Annahme von der Selbsttransparenz des Bewußtseins, von einer starken Form *inneren Wissens*, voraus. Ähnliches gilt für den monologisierenden Gebrauch des Wortes „Ich": Die Wurzel des idealistischen Mißverständnisses liegt in der Unmöglichkeit monologisierender Fehlreferenz, die zu einem Fehlschluß durch einen ungerechtfertigten Einsatz des Existenzquantors führt.[44]

Wenn wir anhand öffentlicher Eigenschaften auf uns selbst Bezug nehmen, dann setzen wir diese publiken Merkmale als Instrumente ein. Wir wollen in einem bestimmten Kontext eine Orientierung vollziehen, indem wir den Sprecher – relativ zu einer bestimmten Beschreibungsebene – lokalisieren. Zumeist ist dies ein *sozialer* Kontext, in dem aus rein pragmatischen Gründen eine Orientierung vorgenommen werden muß: Wenn Thomas Nagel gefragt wird *„Wem gehört der blaue Ford mit den Nummernschildern aus New Jersey, der vor meiner Einfahrt steht?"*, dann wird die Antwort, die er auf diese Frage gibt, wahr oder falsch gemacht durch gewöhnliche Fakten. Diese Fakten befinden sich auf unproblematische und triviale Weise in der Welt. Und sie werden normalerweise in sozialen Zusammenhängen eingesetzt, um Lokalisierungen oder Identifikationen von Personen zu ermöglichen. In allen Fällen interpersoneller Kommunikation stellt die Äußerung von Sätzen des Typs „Ich bin TM" keinerlei transzendentales Mysterium dar. Der Versuch, solche interpersonellen Orientierungsleistungen durch Selbstreferenz zu erbringen, kann jederzeit scheitern, wenn die fraglichen Sätze falsch sind.

Vielen Autoren ist aufgefallen, daß Nagel Begriffe wie „Selbst", „Ich" oder „Bewußtsein" unscharf verwendet. Dies führt zu allerlei subtilen Themenwechseln, manchmal jedoch auch zu Widersprüchen. So hat zum Beispiel Norman Malcom darauf hingewiesen, daß – wenn wir Nagel folgen und das Wesen von Bewußtsein sowie geistigen Zuständen nur im gemeinsamen Auftreten mit der durch den subjektiven *point of view* entstehenden Perspektive sehen – das perspektivelose objektive Selbst zu einem „geistlosen Ding" würde, weil es sich in seinem Streben nach Objektivität so weit vom Standpunkt des *psychologischen* Subjekts entfernt hätte, daß es nun seinerseits unter keinen mentalen Begriff mehr fallen würde. Das führt zu unangenehmen Fragen, wie etwa: *„Wenn das objektive Selbst TN verlassen würde, würde TN dann das Bewußtsein verlieren?"*, oder: *„Wenn es stimmt, daß das Ich TNs ‚Erlebnisse' empfängt*[45]*, wird dann dieselbe Erfahrung zweimal von zwei verschiedenen Subjekten durchlebt?"*[46] Das Grunddi-

[44] Auch Malcolm hat darauf hingewiesen, daß Nagel – ausgehend von einer doppelten Referenz von „Ich" – mit dem objektiven Selbst ein nicht-physisches Individuum erzeugt, weil er „Ich" in beiden Fällen stillschweigend wie einen Designator behandelt und einsetzt. Wenn das so wäre, dann könnte aber auch der Satz „Ich bin Ich" ein *falscher* Satz sein. Vgl. Malcolm 1988: 157, 160; dazu auch Anscombe 1981.

[45] Vgl. Nagel 1992 (1986): 111.

[46] Vgl. Malcolm 1988: 152f.

lemma ist hier eines von Subjekt und Subjektivität, von Abstraktion und Introspektion. Nagel ist bei seinen begrifflichen Analysen dieses Dilemmas häufig in Gefahr, in einen Dualismus abzugleiten, der die beiden Aspekte oder Bewußtseinsfunktionen schlicht *reifiziert.* Ihm ist zudem vorgeworfen worden, daß seine Argumentationen für die Inkompatibilität von subjektiver und objektiver Perspektive auf einer Äquivokation der erkenntnistheoretischen und der metaphysischen Hinsicht auf die subjektiv-objektiv-Dichotomie fußen.[47]

William Lycan hat in seiner Kritik[48] an den perspektivischen *„funny facts"*, die nur aus dem Blickwinkel der ersten Person erfaßt werden können, angenommen, daß jeder dieser Tatsachen eine wahre Aussage entsprechen muß. Da Nagel die logische Struktur der fraglichen Fakten im Dunkeln läßt, geht Lycan davon aus, daß die mysteriöse Perspektivität sich entweder in ihrem Individuen-Element oder in ihrem Eigenschafts-Bestandteil verstecken muß. Das bedeutet: Es gibt (als Bestandteil der eben angesprochenen *Proposition*) einen Individualbegriff oder einen Prädikatausdruck, der nicht auf objektive Weise, also aus der Perspektive der dritten Person erfaßt werden kann – sondern nur aus der des Subjekts. Nun sagt Lycan, daß Individualbegriffe einfach Funktionen von Welten zu Individuen sind und Prädikate Funktionen, die bezüglich einer Welt Mengen von Individuen herausgreifen – nichts an ihnen läßt darauf schließen, daß sie nicht objektiv beschreibbar sein sollten (obwohl es Bedeutungsfunktionen geben mag, die zu komplex für unseren mentalen Apparat sind, d. h. nicht *mental* repräsentiert werden können).

Lycan bemerkt weiterhin, daß Nagel nicht selten essentialistischen Intuitionen durch einen *Akt-Objekt-Jargon* Ausdruck verleiht (einen Vorwurf, den er auch an Kripke richtet[49]). Was heißt das? Es bedeutet, daß wir in der Interpretation mentaler Phänomene und ihres Gehalts deren *Prozeßcharakter* unterschlagen indem wir ein mentales Objekt konstruieren, auf das sich ein subjektiver Akt richtet. In der theoretischen Analyse unseres Erlebnisraums rauben wir damit einem seiner Teile die Ereignishaftigkeit und frieren ihn zu einem *psychischen Einzelding* ein, das scheinbar aus der Ganzheit des inneren Repräsentationsvorgangs gelöst und zu einem diskreten Objekt auf der inneren Bühne werden kann. In der Tat bedient sich Thomas Nagel häufig solcher suggestiven Formulierungen.

> Nagel slips on the same Banana Peel. For example, on p. 164 he talks of taking a „subjective *viewpoint toward* our experience [my italics], and on p. 166 he talks of „how [his experiences] *appear to*" him [ditto]. But experiences are not objects that present appearances to us and toward which we take viewpoints; they

[47] Vgl. Foss 1987: 208 (Fußnote 2).

[48] Vgl. Lycan 1987: Kapitel 7, siehe auch Boër/Lycan 1980, Boër/Lycan 1986.

[49] Vgl. Lycan 1987, Kapitel 2. Lycan hat den hier beschriebenen logischen Ausrutscher in der Analyse von mentalen Zuständen und Ereignissen als mentalen Objekten auf den Namen *banana-peel*-Fehlschluß getauft, weil er ihm so weitverbreitet erscheint, daß er eines eigenen Namens bedarf. Vgl. Lycan 1987: 17.

> *are* appearings and viewpoint-takings. To say, scientistically, that psychologists and brain physiologists are moving toward better and better descriptions of mental facts is not to take the subjectivity out of the experience and replace it with objectivity, but is to push for better description of the experience as a whole, subjectivity and all. The subjectivity is immanent to the experiential event; that does not prevent us from describing the whole event objectively. Not that this is an easy job. Our Functionalist psychologist still has to explain what it is about such-and-such a functional state that makes its proprietor think of it as subjective and point-of-view-y.[50]

Wenn man Thomas Nagels Problem nicht einfach semantisch einplanieren will, könnte man jetzt versuchen, den psychologischen Aspekt des Problems näher zu beleuchten. Man könnte zum Beispiel fragen, ob die notorischen phänomenologischen Reifikationen nicht ihre Wurzeln bereits in repräsentationalen Eigenheiten unseres *Gehirns* haben. Oder man könnte die Frage stellen: Welche zusätzliche Art von Orientierungsleistung könnte mit den fraglichen Sätzen noch erbracht werden, wenn sie in Gestalt des Nagelschen *„philosophical thought"* auftreten? Es könnte ja sein, daß es neben dem Bereich interpersoneller Lokalisierung von Sprechern eine Form interner Selbstreferenz gibt, eine Art *innerpsychische Orientierung des Subjekts*. Der entscheidende Unterschied wäre in einem solchen Falle derjenige zwischen monologisierender sprachlicher und psychischer Selbstbezugnahme – kurz: zwischen linguistischer und mentaler Repräsentation des betreffenden Systems *durch es selbst*. Nach allem was wir bisher über den Prozeß der internen Repräsentation durch biologische Informationsverarbeitungssysteme wissen – und das ist nicht viel –, erfüllt dieser Vorgang eine Funktion *für* das jeweilige System. Vielleicht kann man über die Analyse dieser Funktion zu einem Verständnis von Nagels Problem gelangen, das eine semantische Elimination überflüssig macht.

Der Autor hat selbst versucht, einige Einwände gegen seine Subjektphilosophie vorwegzunehmen, indem er seine Argumente einer semantischen Diagnose unterzogen hat.[51] Er argumentiert dabei gegen eine vorschnelle analytische Trivialisierung des Problems und bietet schließlich eine metaphorische Formulierung für diejenige Wahrheit an, die im Rahmen einer sprachanalytischen Auflösung der Frage seiner Auffassung nach verloren gehen würde:

> Aber nichts davon erklärt oder vertreibt den ganz anders gearteten – philosophischen – Gedanken, den ich hätte, würde ich mir eine Welt voller Menschen ansehen, die Dinge wie „Der Wagen gehört mir" oder „Ich bin seine Frau" von sich geben, und dann zu mir selbst sagen, daß von all den Menschen in dieser zentrumlosen Welt ausgerechnet TN kein anderer ist als *ich: Es ist die Person TN*, durch die nämlich ***dieses denkende Subjekt*** *hier die Welt anschaut.* (Hervorhebung TM)[52]

---

[50] Vgl. Lycan 1987: 79f.
[51] Vgl. Nagel 1992: 103ff, außerdem Ginet/Shoemaker 1983 für weitere Einwände.
[52] Vgl. Nagel 1992: 106.

Bevor man sich der Frage zuwendet, was es heißen könnte, daß *ein denkendes Subjekt die Welt durch eine bestimmte Person betrachtet*, sollte man zwei wichtige Details nicht übersehen.

Erstens geht es hier um die Erklärung der Genese eines *Gedankens*, das heißt um die Aktivierung eines bestimmten *mentalen Repräsentats*. Das Problem ist, wenn wir Nagels Formulierung ernstnehmen, in seinem Kern ein psychologisches – wegen seiner enormen begrifflich-logischen Virulenz jedoch Gegenstand einer *philosophischen* Psychologie. Es ist eine sehr spezielle und problematische Form mentaler Repräsentation, die die Frage aufwirft, *was* in ihr repräsentiert wird. Interpretiert man mentale Repräsentate nach dem Muster propositionaler Repräsentate, dann stellt sich auch die Frage nach der Natur der abgebildeten Sachverhalte. Mit Sicherheit besteht das problematische psychische Phänomen jedoch nicht nur in dem inneren Aussprechen eines Satzanalogs vom Typ „Ich bin TM" – auch wenn manche Philosophen gerne sogar die Phänomene *selbst* noch sprachanalytisch behandeln möchten.

Zweitens tritt die fragliche mentale Repräsentation in einem sehr speziellen Kontext auf: Nämlich immer dann, wenn ich *mit mir selbst* spreche. Ob die von mir zu diesem Zweck eingesetzten Zeichen hochkomplexe neurale Ereignisse in meinem Gehirn (also *natürliche Zeichen*) sind oder ob sie auf der „Makro-Ebene" von Sprechakten anzusiedeln sind, spielt für den gegenwärtigen Zusammenhang vorerst noch keine Rolle. Wichtig ist dagegen, daß wir es mit einem internen Versuch von Orientierung durch Selbstreferenz zu tun haben. Im Gegensatz zu den üblichen, unmysteriösen Formen des Gebrauchs von „Ich" werde ich deswegen ab jetzt von *monologischer Selbstreferenz* sprechen. Es handelt sich hierbei um genau jene Sonderfälle, in denen Sprecher und Adressat identisch sind. Monologische Selbstreferenz ereignet sich immer dann, wenn ich *zu mir selbst* sage, daß ich TM bin.

Betrachten wir vor diesem Hintergrund die interessante, von Thomas Nagel angebotene Metapher etwas genauer. Was kann es heißen, daß *dieses denkende Subjekt die Welt durch die Person TN betrachtet?* Ist damit ein räumliches und doch cartesianisches Bild anvisiert, in dem ein *cogitationes* produzierendes außerweltliches Ich durch ein Fenster im Himmel in das Gehirn von TN blickt?

Angenommen, daß wir die Welt immer nur unter Beschreibungen kennen – zum Beispiel unter wissenschaftlichen oder unter von unserem zentralen Nervensystem erzeugten –, dann geht es für das denkende Subjekt darum, die Welt zu *repräsentieren*. Da wir es mit einem Gedanken zu tun haben, geht es um eine mentale Repräsentation der Welt. Bei dieser Aufgabe soll die Person TNs als *Instrument* eingesetzt werden. Also soll im Rahmen einer psychischen Darstellung der Welt eine Person instrumentalisiert werden, ohne selbst in Form eines wie auch immer gearteten Repräsentats intern gegeben zu sein. Das kann aber nicht so sein: Im Rahmen einer mentalen Weltrepräsentation kann nur das mentale *Repräsentat* einer Person zum kognitiven Werkzeug werden. Außerdem wollen wir ja gerade

verstehen, wie Subjektivität als eine psychologische Eigenschaft entstehen kann. Deswegen dürfen wir sie nicht in Form eines denkenden Subjekts bereits zu Anfang einführen. Wir müssen statt dessen – wenn wir das philosophische Rätsel ernstnehmen und seiner metaphorischen Hülle entkleiden wollen – erklären, wie ein System Subjektivität in Form monologischer Selbstreferenz und durch mentale Repräsentation erzeugen kann. Was die Frage nach irreduziblen Tatsachen der ersten Person angeht, scheint es einen Ausweg zu geben, der mit Tatsachen bezüglich der mentalen Repräsentation von Gehirnen durch sich selbst zu tun hat. Im folgenden Kapitel werde ich versuchen, einen Weg in diese Richtung aufzuweisen. Vorerst sollten wir aber ein letztes Mal zurückkehren zu den beiden anderen auch von Nagel beleuchteten Hauptaspekten des Subjektivitätsproblems, der Subjektzentriertheit und dem qualitativen Gehalt mentaler Zustände.

Warum bilden bestimmte Ereignisse und Zustände in meinem Erlebnisfeld eine Teilmenge von solchen Zuständen, die mir direkt und nicht-derivativ gegeben zu sein scheinen und absolut essentiell als *meine* erscheinen? Es gibt Zustände, die wir häufig als *Zustände unseres Selbst* beschreiben, wenn wir anderen Menschen von ihnen berichten. Die *Meinigkeit* dieser Zustände scheint sie zu bündeln, sie scheint sie zusammenzufassen – und zwar in einem weit stärkeren Sinne als dem einer bloßen Klassenbildung durch Prädikation. In einem unbekannten Vorgang, den ich provisorisch als phänomenale *Reifikation* bezeichnen möchte, wird diese Klasse von Zuständen – so scheint es – zu einem inneren Gegenstand (einem *phänomenalen Individuum*) verdinglicht. Diesen inneren Gegenstand bezeichnen wir nun extern als das (oder als *unser*) „Ich“ oder „Selbst“, und diese Beschreibung erzeugt die bekannten philosophischen Probleme. Der zentrale Fehlschluß der Nagelschen Subjekttheorie scheint daher in der ontologischen Überhöhung eines repräsentationalen Phänomens zu bestehen: Daraus, daß natürliche Repräsentationssysteme wie wir selbst phänomenale Reifikationen durchführen, kann man nicht schließen, daß phänomenale Individuen existieren.[53]

[53] Phänomenale Individuen entstehen z. B. dadurch, daß der *qualitative Gehalt* eines mentalen Zustands (etwa die leuchtende, homogene Farbe eines grünen Nachbildes) als Eigenschaft erster Ordnung (eben eines solchen nicht-physischen Einzeldings) interpretiert werden. Vgl. Lycan 1987: 16ff, 83ff. – Ulrich Blau hat bemerkt, daß eine wichtige Funktion innerer Objekte in der Rettung eines vorkritischen Realitätsmodells besteht: *„Betrachten Sie Ihre phänomenalen Objekte, mentalen Zustände, Erlebnisse oder Sinnesdaten, wie immer sie heißen mögen. Im naiv-realistischen Weltbild erschienen sie dem Subjekt als innere Bilder von äußeren Objekten. Nach dem Verschwinden von Subjekt und Außenwelt hängen sie im Leeren wie Brücken ohne Pfeiler, man versteht sie nicht mehr. Man kann sie kaum individuieren, kaum klassifizieren, kaum beschreiben, nicht einmal vorstellen. Beweis: Schließen Sie die Augen und stellen Sie sich einen grünen Apfel vor. Und nun schließen sie wieder die Augen und stellen sich die* ***Vorstellung*** *eines grünen Apfels vor. Was Sie sich vorgestellt haben, war verkehrt:*
(a) ein grüner Apfel,
(b) ein Kopf, in dem ein grüner Apfel steckt,
(c) ?
(d) eine Mischung von (a)–(c).
Innere Objekte sind Artefakte, die im naiven Weltbild die Kluft zwischen erkennendem Subjekt und erkennbaren äußeren Objekten überbrücken sollen.“ (Blau 1986: 5)

Neuere Forschungsergebnisse zeigen, daß die Subjektgebundenheit mentaler Repräsentate genetisch stark durch bewußtseinsexterne Faktoren bedingt ist. Allerdings sind solche Einsichten erstens demütigend und emotional unattraktiv, zweitens kollidieren sie frontal mit unserem Erleben, mit unserer vorphilosophisch-lebensweltlichen Selbsterfahrung. Wie könnten meine Gedanken und Gefühle jemals[54] etwas anderes sein oder gewesen sein, als *meine* Gedanken und Gefühle? Für die meisten von uns scheint eines sicher: Meine mentalen Zustände sind und werden immer an *mein Ich* gebunden sein, mein Bewußtsein hat ein *Zentrum* und dieses Zentrum kann es nicht verlieren. Es ist diese scheinbare innere Gewißheit, die Nagels Argumenten einen Großteil ihrer Plausibilität verleiht und anti-essentialistischen Einwänden sofort das Handikap der Kontraintuitivität. Ich werde diese dritte Intuition ab jetzt die „Zentriertheits-Intuition" nennen. Sie erzeugt unter anderem das Problem, wie in einem physikalischen Universum phänomenale Standpunkte entstehen können, beziehungsweise: in parallel-distribuiert arbeitenden Gehirnen ein *zentriertes* Bewußtsein.

Daniel Dennett hat darauf hingewiesen, daß jede Theorie, die einen Fortschritt erzielt, anfänglich kontraintuitiv sein wird.[55] Die Zentriertheits-Intuition ist eine spezifisch abendländische Erscheinung und zeigt, wie stark kulturelle Hintergrundannahmen inneres Erleben formen können.[56] Bei näherem Hinsehen zeigt sich jedoch, daß es auch in der westlichen Kultur viele Menschen gibt, denen diese von Nagel ausgebeutete Intuition verlorengegangen ist. Es gibt nämlich viele Menschen, die Epsioden *dezentrierten Bewußtseins* erlebt haben. Solche Episoden können das Ergebnis von Hirntumoren oder Läsionen bestimmter Hirnbereiche sein, von schizophrenen Schüben oder das Resultat der Einnahme psychoaktiver Substanzen. Es gibt sogar Menschen, die sich absichtlich um das Erzeugen nicht-zentrierter Bewußtseinszustände bemühen, etwa durch Meditation oder archaische Ekstasetechniken. Schließlich kennen wir die verschiedenen Fälle von Persönlichkeitsspaltung und von multiplen Persönlichkeiten, welche sich gewissermaßen „ein Gehirn teilen". Wir verfügen über reichhaltiges und nicht mehr bloß anekdotisches Material, das uns die Möglichkeit von nicht-zentriertem, multi-zentriertem und sogar erlebnismäßig fremd-zentriertem Bewußtsein zeigt. Dieses Material demonstriert, daß der fragliche Aspekt der Subjektzentriertheit mentaler Zustände einer Naturalisierung zugänglich ist und daß seine essentialistische Überhöhung keines-

[54] Saul Kripke hat mit einer subtilen Argumentation diesen modalen Intuitionen Ausdruck verliehen; vgl. Kripke 1971, 1972. Er ist darauf häufig mit anti-essentialistischen Argumenten und Verteidigungen der Identitätsthese als einer möglichen *kontingenten* Wahrheit angegriffen worden. Vgl. Bartels 1984, Lycan 1987 (Kapitel 2), McGinn 1980 (bzw. 1981), Metzinger 1985: 213ff; ferner Sher 1977, Leplin 1979, Levin 1979, 1985, Hill 1981, McGinn 1981, McMullen 1985.

[55] Vgl. Dennett 1987: 6.

[56] Vgl. etwa Taylor 1987 und die „konnektionistische" Antwort auf den Vorwurf der irreduziblen Plastizität und kulturellen Einbettung des menschlichen Bewußtseins von Churchland 1989: 129ff.

falls zur zwangsläufigen – weil einzig verfügbaren – theoretischen Strategie zu werden braucht.

Wenden wir uns abschliessend dem ersten von Nagel aufgegriffenen Problemaspekt zu: Subjektivität qua *phänomenaler Gehalt*. Wie ist es, Batman zu sein? Angenommen, wir würden bereits heute über die Gesamtheit möglicher wissenschaftlicher Erkenntnisse bezüglich der mentalen Zustände von Batman verfügen – was genau würden wir dann *nicht* wissen? Wir wüßten – so Nagel – nicht, *wie es sich anfühlt*, als Erlebnissubjekt durch Batmans innere Zustände hindurchzugehen. Vielleicht wären wir in der Lage, eine erschöpfende Beschreibung des Superhelden von der Biologie über die Neuroinformatik bis hinauf zur Psychologie anzubieten – aber selbstverständlich wüßten wir nicht, *wie es ist*. Das Problem dieser und verwandter antinaturalistischer Denkfiguren[57] liegt darin, daß sie dem Nicht-Skeptiker keine genaue Beschreibung des Explanandums anbieten können. Es ist nicht klar, *wovon* überhaupt gesagt wird, daß es sich dem objektivierenden Zugriff der Wissenschaft prinzipiell entzieht.[58] Das Problem des phänomenalen Gehalts mentaler Zustände ist – häufig in vorschneller Abgrenzung von ihrem intentionalen Gehalt – in der neueren Diskussion überwiegend unter dem Stichwort „Qualia" abgehandelt worden.[59] Wenn man jetzt unter der Hinsicht phänomenalen Gehalts „Wie es ist" als *die Gesamtheit der zu einem gegebenen Zeitpunkt auftretenden Qualia* interpretiert, dann sieht man sofort die Schwäche skeptischer Subjektivitäts-Argumente der Nagelschen Art. Denn Qualia sind dem üblichen Verständnis zufolge singuläre, homogene Aspekte mentaler Zustände, deren Gehalt sprachlich nicht adäquat repräsentiert werden kann und die deshalb weitgehend inkommunikabel sind: Wir können einem Blindgeborenen nicht erklären, was der Kern eines Türkis-Erlebnisses ist – und Superman wird nie verstehen, wie es ist, Batman zu sein. Denn es handelt sich beim Auftreten von Qualia um präreflexive innere Phänomene, die vor und unabhängig von jeder sprachlichen Kategorisierung in erlebnismäßig

[57] Vgl. auch Jackson 1982.

[58] Kathy Wilkes hat wiederholt (1984, 1988) dafür argumentiert, daß der alltagssprachlich-phänomenologische Begriff „Bewußtsein" keine natürliche Art herausgreift und aus diesem und anderen Gründen für wissenschaftlich orientierte Theorien des Geistes grundsätzlich irrelevant ist. *„One of the pressing problems for the brain and behavioural sciences is the construction of a taxonomy of explananda. It is evident that pain, under some description, must feature in such a taxonomy. However, what counts as pain for the purposes of the sciences may prove to be in part at least a product of what kinds of phenomena the explanans serves to unify, and thus may diverge from the extension of the class of phenomena picked out by the vernacular term. . . . The adaptation that the term ‚consciousness' would need to undergo before it could be made to cover tidily a systematically-related bunch of behaviours would be so great that a study of this ‚consciousness' would no more be a study of consciousness as we think of it than the study of the spin of an electron can inform us about the behaviour of whippings-tops."* (Wilkes 1984: 234, 239)

[59] Vgl. etwa Block 1980, Churchland 1981b, 1985b, 1989, Davis 1982, Dennett 1978, 1979, 1981, 1988, Jackson 1982, 1986, Kitcher 1979, Lewis 1980, 1990, Lycan 1973, 1981a, 1987, Shoemaker 1975, 1981, 1982.

sehr direkter und konkreter Weise gegeben sind. Wenn es aber wirklich Aspekte unserer phänomenalen Realität gibt, die unaussprechlich sind, dann birgt eben jene Eigenschaft die Gefahr in sich, als Prämisse skeptische Argumente in die Nähe des Obskurantismus zu rücken.

Wovon man nicht sprechen kann, damit kann man nämlich auch nicht für die prinzipielle Unvollständigkeit der wissenschaftlichen Kosmologie argumentieren. Wenn es Elemente unseres Innenlebens und unserer Sinneserfahrung gibt, deren Gehalt sich der sprachlichen Beschreibung entzieht, dann kann man den Gehalt dieser Elemente auch nicht in wissenschaftliche oder philosophische Argumente einbringen, weil er *ex hypothesi* nicht theoriefähig ist. Will man unaussprechliche Erlebnisqualitäten als Waffe gegen monistisch-reduktionistische Theorien des Mentalen einsetzen, dann benötigt man die starke cartesianische Zusatzprämisse der vollständigen Selbsttransparenz des Bewußtseins. *Weiß* ich – in einem nicht-trivalen, philosophisch interessanten Sinn von „innerem Wissen" – überhaupt, *wie es ist*, ich selbst zu sein? Habe ich wirklich einen nicht-sprachlichen epistemischen Zugang zu inneren Phänomenen? Eine der zentralen Schwächen von Nagels Fledermaus-Argument liegt in dem Umstand, daß es auf der starken Prämisse fußt, daß Introspektion und subjektives Erleben epistemische Vorgänge sind, durch die eine starke Art *privaten Wissens* geschaffen wird.[60] Die Annahme dieser Möglichkeit führt fast zwangsläufig zu einer Postulierung nicht-physikalischer, privater Objekte von Wissen und verbaut so jede Möglichkeit eines weiteren theoretischen Fortschritts.

Die Stärke von Nagels Betonung des qualitativen Gehalts unserer inneren Erfahrung liegt in ihrer direkten intuitiven Plausibilität und darin, daß sie phänomenales Bewußtsein als Thema reetabliert. Ihre Schwäche liegt in einem zu weit gefaßten bzw. ungeklärten, aber immer vorausgesetzten Begriff inneren Wissens. In der Verblüffung darüber, daß es hier einen Problemaspekt gibt, der resistent gegen jede Art physikalistischer oder funktionalistischer Analyse zu sein scheint, kann man leicht übersehen, daß dieser Aspekt niemals wirklich genau beschrieben worden ist. Bevor Schmerzerlebnisse und Rotempfindungen zu Explananda zukünftiger Theorien des Geistes werden können, müssen wir besser verstehen, *was* an inneren Episoden dieser Art es uns unmöglich macht, sie als das Zusammenwirken von Elementen einer tiefer liegenden Beschreibungsebene zu verstehen. Solange

[60] Das gilt auch für das „Knowledge-Argument" von Frank Jackson. Was ist das Verhältnis von sprachlichen und nicht-sprachlichen Erkenntnisansprüchen in bezug auf *subjektive* Zustände? Eine nominalistische Konzeption der inneren Erfahrung gerät in folgendes Trilemma: Wenn innere Erfahrung als epistemisches Phänomen immer propositional ist und es eine innere nicht-propositionale Erfahrung von mentalen Zuständen gibt, dann kann dies keine Form von *Wissen* sein: Unser eigenes Bewußtsein ist opak. Wenn Erfahrung immer propositional ist und innere Erfahrung ein epistemisches Phänomen, dann *muß* die innere Erfahrung propositional sein. Wenn es dagegen eine innere, nicht-propositionale Erfahrung gibt, die gleichzeitig ein *epistemisches Phänomen* darstellt, dann ist die nominalistische Konzeption von Erfahrung verfehlt. Eine gute Exposition dieses Problems gibt Peter Bieri in *„Nominalismus und innere Erfahrung"*; vgl. Bieri 1982.

dies nicht geleistet ist, steht ein Erkenntnispessimismus bezüglich des qualitativen Gehalts phänomenaler Repräsentate auf schwachen Füßen.

Die Schwierigkeiten der Nagelschen Subjektphilosophie sind nicht die Schwierigkeiten der Freunde des metaphysischen Subjekts, sondern die Probleme eines rationalistischen Unternehmens. Dieses Unternehmen besteht unter anderem darin, die Dimension des Subjektiven in einer Zeit rasanten Erkenntniszuwachses gegen reduktionistische Euphorie und utopistisches Theoretisieren zu verteidigen, sowie unsere eigene Innerlichkeit und die traditionell mit ihr verknüpften philosophischen Rätsel weiter ernst zu nehmen. Wie die Erfolgschancen eines solche Unternehmens einzuschätzen sind – und ob es naturalistische Alternativen zu ihm gibt – können wir jetzt besser beurteilen, indem wir die Fragen des ersten Kapitels aus zwei Richtungen beantworten: Unter Rekurs auf einen der wichtigsten antinaturalistischen Autoren der Gegenwart und in allgemein-systematischer Hinsicht.

5. Kapitel

Vom Subjekt zum Selbstmodell: Perspektivität ohne Ego

5.1 Der totale Flugsimulator und der kleine rote Pfeil

Welches Bild des menschlichen Geistes ergibt sich aus den bisher angestellten Überlegungen? Bevor wir uns den spezifisch philosophischen Fragen zuwenden, möchte ich eine *technische* und eine *repräsentationale* Metapher anbieten, die einen Teil der den vorangegangenen Bemerkungen zugrundeliegenden Gedanken veranschaulichen sollen. Die erste dieser Metaphern ist der *Flugsimulator*. Ein Flugsimulator ist ein Gerät, an dem zukünftige Piloten ausgebildet werden. Es dient auch dazu, das Verhalten in unvorhergesehenen und kritischen Situationen zu üben, ohne das Risiko eines tatsächlichen Absturzes einzugehen. Flugsimulatoren wurden bereits zu Anfang des Jahrhunderts eingesetzt, seitdem sind sie technisch ständig verbessert worden. In den heute technisch bereits überholten Standardmodellen befinden sich die Kandidaten häufig in einer Kabine, die auf großen Teleskopfüßen oder einer Bewegungsplattform ruht. Diese Teleskopfüße werden von einem Rechner angesteuert, der auf diese Weise für den in der Kabine sitzenden Flugschüler alle Bewegungen eines wirklichen Flugzeug nachahmen kann. In der Kabine befindet sich ein realistisch gestaltetes Cockpit mit allen Instrumenten und Steuerungswerkzeugen, die ein echtes Flugzeug auch besitzt. Der Schüler blickt auf den ebenfalls von einem Computer angesteuerten Videobildschirm, der ihm eine visuelle Simulation des Blicks aus dem Cockpit liefert. In fortgeschrittenen Modellen ist dieser Bildschirm bereits durch einen Datenhelm ersetzt, der über zwei räumlich versetzte Kleinbildschirme den Ausblick in eine dreidimensionale Rundumgrafik ermöglicht. Sie zeichnet sich durch eine „Unendlichkeits-Optik“ aus: Durch eine spezielle Programmiertechnik befindet sich der virtuelle Fokus des Bildes in mehr als zehn Metern Entfernung. Wenn der Proband „aus dem Fenster“ schaut, kann er seinen Blick auch auf weit entfernte Objekte fokussieren – obwohl das tatsächliche Bild sich nur wenige Zentimeter von seinem Gesicht entfernt befindet. Diese visuelle Simulation der Außenwelt wird mit großer Geschwindigkeit und in Abhängigkeit von den Handlungen des Piloten ständig aktualisiert. Auch der propriozeptive und der kinästhetische Sinn lassen sich heute stimulieren, zum Beispiel durch einen *Seat Shaker*, der vom Strömungsabriß bei kritischen Geschwindigkeiten bis zu den Vibrationen des Nachbrenners eine ganze Reihe körperlicher Empfindungen des Piloten zu simulieren hilft. Auf diese Weise kann ein Flugschüler den Umgang mit den Bordinstrumenten sowie die Reaktion eines Luftfahrzeugs auf seine Handlungen kennen-

lernen und gefahrlos die wichtigsten Grundoperationen einüben, deren Beherrschung für einen guten Piloten unerläßlich ist.

Menschliche Gehirne funktionieren auf sehr ähnliche Weise. Aus gespeicherten Informationen und dem ständigen Input, den ihnen die Sinnesorgane liefern, konstruieren sie ein internes Modell der äußeren Wirklichkeit. Dieses Modell ist ein *Echtzeit-Modell*: Es wird mit so hoher Geschwindigkeit und Effektivität aktualisiert, daß wir es im allgemeinen nicht mehr *als* ein Modell erleben. Die phänomenale Wirklichkeit ist für uns kein von einem Gehirn erzeugter Simulationsraum, sondern auf sehr direkte und erlebnismäßig unhintergehbare Weise schlicht *die Welt, in der wir leben*. Einen Flugsimulator dagegen erkennen wir auch dann, wenn wir als Flugschüler gerade konzentriert mit ihm arbeiten, immer noch *als* Flugsimulator – wir glauben niemals, daß wir *wirklich* fliegen. Das liegt daran, daß das Gehirn uns ein wesentlich besseres Modell der Welt liefert als der Computer, der den Flugsimulator ansteuert. Die Bilder, die unser visueller Cortex erzeugt, sind wesentlich schneller, zuverlässiger und besitzen eine viel höhere Auflösung und einen größeren Detailreichtum als die Bilder auf dem Monitor des Übungssimulators. Darum erkennen wir die Bilder auf dem Monitor auch jederzeit *als* Bilder, weil wir einen wesentlich höheren repräsentationalen Standard besitzen, mit dem wir sie vergleichen können. Wenn die Teleskopfüße die Kabine, in der der Flugschüler seine Übungsstunde absolviert, rütteln oder stoßen, um das Durchfliegen von „Luftlöchern“ oder die Konsequenzen ungeschickter Steuermanöver zu simulieren, dann werden auch diese Rüttel- und Stoßbewegungen uns nicht wirklich täuschen können. Denn die auf unseren propriozeptiven und kinästhetischen Körperwahrnehmungen beruhenden mentalen Modelle unserer *eigenen* Körperbewegungen sind viel detailreicher und überzeugender, als die von einem Rechner erzeugten Simulationen von Bewegungen des Flugzeugs es jemals sein könnten.

Unser Gehirn unterscheidet sich aber von einem Flugsimulator noch in vielen anderen Punkten. Es verfügt über wesentlich mehr *Modalitäten*: Das Sehen, die akustische Wahrnehmung, den Geruchs- und Geschmackssinn, den Tastsinn und die „Eigenwahrnehmung“ des Körpers. Es ist in der Lage, die aus diesen verschiedenen Modalitäten stammende Information bruchlos zu einem *einheitlichen* Modell der Wirklichkeit zu verschmelzen (eine Aufgabe, die auch im Flugsimulator noch dem Gehirn des Probanden überlassen bleibt). Es arbeitet zudem wesentlich schneller. Die von ihm erzeugten multimodalen Bilder der Wirklichkeit sind zuverlässiger und detailreicher als die künstlichen Bilder, die wir heutzutage kennen – eine Situation die sich recht bald ändern wird.[1] Außerdem sind Gehirne im

[1] Es gibt mittlerweile interaktive Systeme, die wesentlich umfassendere *virtuelle Realitäten* erzeugen als Flugsimulatoren. Diese virtuellen Realitäten enthalten bereits ein dreidimensionales visuelles Modell der simulierten Welt und auch schon rudimentäre *Selbstmodelle* (deren Input durch einen von der Versuchsperson getragenen Datenhandschuh oder -overall erzeugt wird). Diese rudimentären *virtuellen Selbste* (deren „programmiertechnischer Urahn“ der

Gegensatz zu Flugsimulatoren nicht auf einen eng umgrenzten Anwendungsbereich fixiert, sondern offen für eine Unendlichkeit von repräsentationalen Situationen und Simulationsproblemen. Gehirne sind *General Problem Solvers*.[2] Das in unserem Zusammenhang wichtigste Unterscheidungsmerkmal zwischen einem menschlichen Gehirn und einem Flugsimulator ist jedoch ein ganz anderes: *Menschliche Gehirne simulieren den Piloten gleich mit*.

Denn natürlich gibt es keinen Homunkulus im System. Es gibt aber die Notwendigkeit für das System *als Ganzes*, sich seine eigenen inneren und äußeren Handlungen selbst zu erklären. Es muß nämlich ein repräsentationales Werkzeug besitzen, mit dessen Hilfe es kritische Eigenschaften seiner selbst durch interne Simulation überwachen und sich selbst die Geschichte seiner eigenen Handlungen auch intern als *seine* Geschichte darstellen kann. Dieses Werkzeug ist das, was ich als das mentale Selbstmodell des Organismus bezeichnet und im dritten Kapitel genauer beschrieben habe. Das Gehirn unterscheidet sich nun von einem Flugsimulator unter anderem dadurch, daß es nicht von einem Piloten benutzt wird, der vorübergehend in es „eingestiegen" ist. Es operiert wie ein *„Totaler Flugsimulator"*: Der *Totale Flugsimulator* ist ein selbstmodellierendes Flugzeug, das schon immer ohne Piloten fliegt und in *seinem* Flugsimulator ein komplexes Bild von sich selbst erzeugt. Weil es aber in einem naiv-realistischen Mißverständnis dieses Bild als *Ding* interpretiert, entsteht in seinem Flugsimulator „der Pilot". Dieses repräsentationale Mißverständnis erzeugt gleichzeitig auf der Ebene des subjektiven Erlebens ein phänomenales *Selbst*mißverständnis. Jenes Defizit an subjektivem Wissen über die Entstehungsbedingungen und die innere Struktur unseres Selbstbewußtseins zieht dann das platonische Bild vom Steuermann und die Geburt des cartesianischen Mythos nach sich, die vielen falschen Theorien vom „Piloten", der vorübergehend in den Körper „eingestiegen" ist. All diese Theorien erweisen sich nun im Rahmen unseres objektiven Wissenszuwachses auf schmerzliche Weise als unwahr. Im Gegenteil: Das Gehirn *aktiviert* den Piloten, und zwar immer dann, wenn es ihn als repräsentationales Werkzeug benötigt, um die

---

Mauszeiger ist) sind jedoch noch nicht *benutzerfixiert*, weil sie nicht funktional untrennbar mit der Person verbunden sind, die den Datenoverall trägt. Aus diesem Grund werden wir auch niemals wirklich „durch den Bildschirm hindurch" in den *Cyberspace* eintreten können. Die zeitgenössische Begeisterung für das Vordringen des Menschen in künstliche virtuelle Welten übersieht, daß wir uns immer schon in einem *biologisch* generierten *„Phenospace"* befinden: Innerhalb einer durch mentale Simulation erzeugten virtuellen Realität. Trotzdem ist die technologische Metapher des *Cyberspace* ein wichtiger Fingerzeig, weil sie uns interessante Intuitionen bezüglich unserer eigenen phänomenalen Zustände liefern kann. Künstliche Systeme, die in Echtzeit interaktive virtuelle Realitäten erzeugen, geben uns ein erstes Gefühl dafür, wie aus purer Informationsverarbeitung komplette Erlebniswelten entstehen können.

[2] Ein *GPS* (General Problem Solver) im Sinne des klassischen KI-Ansatzes ist ein System, das sich mittels einer „Zweck-Mittel-Analyse" *(„means-end-analysis")* stufenweise durch einen Raum möglicher Problemlösungsstrategien bewegt. Das Grundverfahren ist dabei von *speziellen* Heuristiken und Operatoren unabhängig, deshalb leicht auf viele verschiedene Problemklassen anwendbar. Vgl. Haugeland 1987: 155ff.

Aktivitäten des Gesamtsystems zu überwachen und mental abzubilden. Braucht das System für einen gewissen Zeitraum kein funktional aktives Selbstmodell mehr, so wird es einfach abgeschaltet. Mit dem Selbstmodell verschwindet auch das Erlebnissubjekt: Der Schlaf ist der kleine Bruder des Todes.

Wenden wir uns der *zweiten*, der *repräsentationalen* Metapher für die Struktur unseres phänomenalen Bewußtseins zu. Menschliche Organismen im Wachzustand gehören zu einer bestimmten Klasse informationsverarbeitender Systeme, nämlich zur Klasse der *Selbstmodellgeneratoren*. Von den Angehörigen anderer Systemklassen unterscheidet Selbstmodellgeneratoren die Fähigkeit, die intern von ihnen erzeugten Repräsentationsräume durch ein Selbstmodell zu ergänzen. Dadurch werden diese Räume zu zentrierten Repräsentationsräumen: Sie gleichen jetzt einer *fixierten* inneren Landkarte der Welt, die auf die Interessen und Bedürfnisse eines individuellen Benutzers zugeschnitten ist – ähnlich wie der an der Wand eines U-Bahnhofes fest angebrachte Stadtplan mit einem kleinen roten Pfeil und dem Hinweis „SIE BEFINDEN SICH HIER“. Dieser kleine rote Pfeil ist das „Selbstmodell des Stadtplanbenutzers“, das die Position und damit auch die Interessen möglicher Benutzer eines solchen externen Repräsentats in diesem noch einmal spezifiziert. Durch den kleinen roten Pfeil und den indexikalischen Hinweissatz „SIE BEFINDEN SICH HIER“ verliert der Stadtplan seine Universalität und wird zu einem Orientierungswerkzeug, das nur noch an einem einzigen Ort in der Welt erfolgreich von allen potentiellen Benutzern eingesetzt werden kann.

Die von menschlichen Gehirnen erzeugten multimodalen Landkarten der Welt sind dagegen *generelle* Realitätsmodelle, die sich der jeweiligen Situation des Organismus anpassen und in Echtzeit aktualisiert werden. Da sie zudem *interne* Modelle der Welt sind, ist der Benutzer, dem sie dienen, in allen Situationen faktisch derselbe. Im Gegensatz zu fest installierten Stadtplänen in U-Bahnhöfen ist nicht die Anwendungssituation fixiert und die Benutzerklasse variabel, sondern das *System* über alle repräsentationalen Situationen hinweg identisch, während die Problemklasse eine sehr allgemeine ist. Wenn man so will, dann sind Selbstmodelle die kleinen roten Pfeile, die in komplexen mentalen Landkarten der Wirklichkeit die Eigenschaften des mentalen Geographen selbst *für* ihn noch einmal abbilden.[3] Deshalb verwandeln sie – solange sie funktional aktiv sind – die Realitätsmodelle, in die sie vom System eingebettet werden, in *benutzerzentrierte* Repräsentate: Nicht nur aus Gründen ihrer physikalischen Inter-

[3] Dieser Vergleich hinkt unter anderem deshalb, weil mentale Modelle keine Variablen besitzen (vgl. Abschnitt 2.2): Das bewußte Selbstmodell ist immer interpretiert. Der kleine rote Pfeil auf dem Stadtplan an der Wand des U-Bahnhofes dagegen ist eine Variable, weil verschiedene Passagiere die Karte benutzen können, indem sie sich mit dem kleinen roten Pfeil identifizieren (sich sozusagen in das Realitätsmodell des Stadtplans „inkarnieren“).

nalität[4], sondern auch durch ihre strukturell-repräsentationale Fixierung auf einen einzigen Anwender, sind auf diese Weise zentrierte Realitätsmodelle nur noch für ein einziges System sinnvolle Instrumente. Abstrakte Organe wie mentale Modelle der Welt und des Selbst sind deshalb auch nicht transplantierbar, denn ihr *funktionales Profil* – das durch extraorganismische Relationen geprägt wird – kann nicht beliebig in ein anderes System überführt werden.[5] Die Einzigartigkeit jedes phänomenalen Subjekts hat somit ihr Gegenstück in der Einzigartigkeit der funktionalen Eigenschaften des ihm zugrundeliegenden Selbstmodells. Dieses Selbstmodell ist der kleine rote Pfeil, den ein menschliches Gehirn benutzt, um sich in der von ihm aufgebauten inneren Simulation der Welt zu orientieren.

Das Bild vom totalen Flugsimulator und dem kleinen roten Pfeil ist ein allgemeines intuitives Bild, das sich aus den Überlegungen der vorangegangenen Kapitel ergibt. Es sollte den Grundgedanken der vorangegangen Untersuchungen noch einmal kurz illustrieren. Mit diesem Bild und den Überlegungen, die zu ihm geführt haben, im Kopf lassen sich nun Antworten entwickeln in bezug auf den im letzten Abschnitt des ersten Kapitels formulierten Minimalkatalog von Fragen in Zusammenhang mit dem philosophischen Problem der Subjektivität mentaler Zustände. Ich hatte diesen Katalog in **logisch-semantische**, **erkenntnistheoretische** und **psychologische** Fragen gegliedert. Da das Interesse dieser Arbeit sich primär auf psychologische Subjektivität richtet, werde ich mich diesen Fragen nun auch zuerst zuwenden.

5.2 Subjektivität in zentrierten Repräsentationsräumen: *Vorläufige Antworten*

Informationsverarbeitende Systeme können psychologische Eigenschaften instantiieren, indem sie ganz bestimmte funktionale und repräsentationale

[4] Diese Eigenschaft könnten sie prinzipiell verlieren, etwa indem sie dupliziert und in *anderen* Systemen aktiviert werden. Wenn man durch irgendeine zukünftige Technologie der funktionalen Klonierung in meinem Gehirn das Realitätsmodell einer Fledermaus aktivieren könnte, würden zwar phänomenale Fledermauszustände entstehen. Andererseits wären solche Zustände aber im biologisch-physikalischen Kontext *meines* Organismus hochgradig afunktionale Simulate, deren intentionaler Gehalt gegen Null ginge. „Transplantierte Selbstmodelle" stünden zu dem sie nun beherbergenden Organismus nur in einer *schwachen Ähnlichkeitsrelation*.

[5] Wenn es stimmt, daß die subjektiven Zustände eines physikalischen Systems höherstufige funktionale Zustände desselben sind, dann nehmen diese Zustände auch einen bestimmten Ort in der kausalen Matrix der Welt ein. Dieser Ort wird durch interne, physikalische Zustände des Systems realisiert und damit fixiert. *„Subjective facts accessible from other points of view will remain opaque to us, not because there is some special picture which must remain hidden from us or entitites which we can never glimpse, but because we can not bring our causal position inside another self conscious subject to duplicate his own location."* Van Gulick 1985: 68.

Zustände in sich erzeugen. Alle diese Zustände besitzen physikalische Beschreibungen. Die für Verhaltenserklärungen und die Zuschreibung psychologischer Eigenschaften interessanten Charakteristika dieser intern erzeugten Systemzustände erfassen wir aber auf wesentlich abstrakteren, das heißt *höher* angesiedelten Beschreibungsebenen. Das tun wir, indem wir sie über ihre kausale Rolle in der globalen Ökonomie des Systems individuieren oder ihnen intentionalen und phänomenalen Gehalt zuschreiben. Die Rückbindung dieser abstrakteren Beschreibungssysteme an die sich aus empirischen Erkenntnissen ergebende, in ständigem Fluß befindliche *wissenschaftliche* Taxonomie mentaler Zustände einerseits und an unsere christlich-cartesianisch geprägte Alltagsphänomenologie andererseits erzeugt die bekannten begrifflichen Dissonanzen und philosophischen Probleme.

Ich werde nun zu zeigen versuchen, welche Antworten man vor dem Hintergrund einer naturalistischen Theorie mentaler Repräsentation auf die eingangs formulierten Fragen geben kann. Da der zugrundegelegte begriffliche Rahmen der mentalen Modellierung bewußt plastisch angelegt ist, also offen für zukünftige Erweiterungen, semantische Anreicherungen und Eliminationen sein soll, sind alle diese Antworten vorläufige Antworten: Der Leser sollte sie nicht als Endstationen und fixe Positionen betrachten, sondern eher als *Türen*. Hinter diesen Türen liegen neue Räume postmetaphysischer Selbsterkenntnis, die uns sowohl mit einem neuen Verständnis unserer selbst als geistigen Wesen als auch mit einer Vielzahl neuer Probleme konfrontieren werden.

Die erste Gruppe von Problemen besteht nicht aus theoretischen, sondern aus *intuitiven* Problemen: Das eben skizzierte Bild vom Gehirn als totalem Flugsimulator, der unter anderem auch *uns selbst* erzeugt, kollidiert frontal mit unserem Selbsterleben. Selbst wenn wir **SMT**, die Selbstmodell-Theorie der Subjektivität, für akzeptabel halten, können wir eigentlich nicht wirklich *glauben*, daß wir als phänomenale Wesen die von „unseren" Gehirnen aktivierten Selbstmodelle sind.

> **P 1**: *Wie entstehen cartesianische Intuitionen – wie z. B. die Kontingenz-Intuition, die Unteilbarkeits-Intuition, das Gefühl der direkten Gegebenheit mentaler Inhalte?*

Unsere Intuitionen sind eine direkte Widerspiegelung von „impliziten Annahmen" über das Wesen der Wirklichkeit, die in unserem mentalen Realitätsmodell kodiert sind. Diese „impliziten Annahmen" sind jedoch keine verborgenen satzartigen Strukturen, die den Gehalt unserer bewußten kognitiven Operationen beeinflussen, sondern letztlich Eigenschaften der *funktionalen Architektur* unseres Nervensystems. Diese funktionale Architektur ist das Resultat eines erbarmungslosen, Millionen Jahre andauernden biologischen Optimierungsvorganges. Die funktionale Adäquatheit innerer Operationen relativ zu einer gegebenen Umwelt allein stellt aber

noch keine *epistemische Rechtfertigung* der durch diese Operationen aktivierten mentalen Modelle dar – eine solche These wäre nur eine weitere Variante des genetischen Fehlschlusses.

Allerdings verkörpert die funktionale Architektur biologischer Gehirne im Normalfall eine große Menge an Information darüber, was in der Umwelt des Systems und seiner Vorfahren *möglich* ist. Wenn wir uns ein konnektionistisches System – etwa ein neuronales Netz – als ein durch eine sehr große Anzahl von instanzbasierten Lernschritten intern konfiguriertes System denken, dann fällt auf, daß ein solches System eine innere Energielandschaft in sich erzeugt hat, die manche seiner Aktivierungszustände zu wahrscheinlicheren und andere zu unwahrscheinlicheren macht. Für die Wahrscheinlichkeit, mit der ein System sich relativ zu einem gegebenen Input in bestimmten Partitionen seines Zustandsraums aufhalten wird, beziehungsweise für die notwendige Energie, die aufgewendet werden muß, um einen bestimmten Aktivierungszustand zu stabilisieren, gibt es präzise mathematische und physikalische Modelle. Die technischen Details sind für die philosophische Fragestellung weniger interessant als die Einsicht, daß die funktionale Mikroarchitektur eines solchen Systems festlegt, welche *mentalen Simulationen* ihm möglich sind und welche nicht. Mit anderen Worten: Die biologische Geschichte unserer Gehirne in ihren speziellen Umwelten bestimmt *was wir uns vorstellen können* – und auch, was wir uns *nicht* vorstellen können. Wir können uns das Kontinuum der Raumzeit nicht vorstellen, weil das drei räumliche und eine distinkte, unidirektionale zeitliche Dimension beinhaltende mentale Realitätsmodell, das unsere Gehirne im Wachzustand aktivieren, für *unsere* biologische Umwelt ausreichend war. Das heißt: Es greift die für eine funktionale Adäquatheit unseres Verhaltensprofils wichtigen Eigenschaften der physikalischen Welt zuverlässig heraus. Wir können uns den dreizehndimensionalen Schatten eines vierzehndimensionalen Würfels nicht vorstellen, weil unser visueller Cortex niemals mit solchen Objekten konfrontiert war und deshalb die repräsentationalen Ressourcen zur Durchführung der entsprechenden mentalen Simulation nicht aufgebaut hat, beziehungsweise weil die ihr korrespondierenden Aktivierungszustände in einem extrem weit vom Harmoniemaximum entfernten Bereich seines Zustandsraums liegen. Über die Struktur der Welt im allgemeinen sagt dieses Faktum brutum nichts aus.

Die *Unteilbarkeits-Intuition*, der Descartes in seiner *Sechsten Meditation* den Charakter einer unmittelbar evidenten Wahrheit bezüglich jeder *res cogitans* zu verleihen sucht, hat ihre Wurzeln in unserer Unfähigkeit, unseren inneren Repräsentationsraum oder das in ihn eingebettete Selbstmodell im Rahmen einer mentalen Simulation zu *spalten*: Als physische Systeme sind wir aufgrund von Eigenheiten unserer funktionalen Architektur in Standardsituationen nicht in der Lage, mehrere stabile und konsistente mentale Realitäts- und Selbstmodelle in uns zu aktivieren. Da die meisten von uns noch nie hypnotisch induzierte Dissoziationen, psychotische Schübe oder andere der von mir in den Abschnitten 2.3.2 und 3.2.2 diskutierten Beispiele für phänomenale Spaltungszustände erlebt haben, kennen

wir die entsprechenden repräsentationalen Gesamtzustände nicht und können sie uns auch nicht *vorstellen*. Die Tatsache, daß die meisten von uns gewisse mentale Simulationen nicht durchführen können, sagt aber nichts über den analytischen Status der Sätze aus, mit denen wir uns extern auf die entsprechenden mentalen Modelle beziehen könnten. Das gilt auch mit Blick auf die von Thomas Nagel so eindrücklich explizierte *Kontingenz-Intuition*, die uns glauben läßt: „Ich hätte auch ein(e) ganz andere(r) sein können!" Der von Vertretern eines essentialistischen Subjektbegriffs immer wieder nahegelegte Fehlschluß beruht nämlich auf einer Verwechslung von mentaler Simulation oder repräsentationalem Gesamtzustand einerseits und philosophischer Ontologie andererseits. Die unbestrittene Tatsache ist, daß wir alle uns sehr wohl *vorstellen* können, ganz andere öffentliche und private Eigenschaften zu besitzen – etwa die von Immanuel Kant. Das heißt aber nichts anderes als das Folgende: Unsere Gehirne können in einer bestimmten Partition des von ihnen geöffneten phänomenalen Raumes ein fiktives *Selbstsimulat* aktivieren, das die uns bekannten Eigenschaften von Immanuel Kant mehr oder weniger gut abbildet. Um den entsprechenden fiktiven Bewußtseinzustand zu erzeugen, müßte allerdings das alte Selbst vollkommen verloren gehen und ein entsprechender *repräsentationaler Gesamtzustand* aktiviert werden. Wir wären dann Systeme, die aufgrund eines hochgradig afunktionalen Selbstsimulats *glauben*, sie seien Immanuel Kant und sich selbst auch entsprechend *erleben*. Wie die meisten meiner Kollegen wissen, treten solche Systeme tatsächlich immer wieder einmal auf. Aber gerade solche wahnartigen, auf außer Kontrolle geratenen mentalen Simulationsversuchen beruhenden Zustände zeigen ja, daß es eben *keinen* essentiellen Wesenskern des phänomenalen Ich mehr gibt. Aus der intendierten Erzeugung von Selbstsimulaten dagegen kann man keine Schlüsse auf die ontologische Struktur der Welt oder die Analytizität der diese Simulate extern indizierenden Sätze ziehen: *Repräsentationale* Möglichkeiten rechtfertigen weder Existenzannahmen noch Aussagen über *logisch* mögliche Welten (auf das, was Nagel „den Übergang von einem objektiven in ein subjektives Weltbild" nennt, gehe ich weiter unten ein). Was in allen subjektiven repräsentationalen Gesamtzuständen – auch dann, wenn sie zusätzliche simulierte Selbstmodelle enthalten – gleich bleibt, ist ihre Zentriertheit. Die *phänomenale Perspektivität* des entsprechenden Bewußtseinszustands ist aber kein *perspectival fact* im ontologisierten Nagelschen Sinne, sondern einfach eine psychologische Eigenschaft, die von einem System durch die Aktivierung eines zentrierten Realitätsmodells instantiiert wird.

Werfen wir abschließend auch einen kurzen Blick auf die mythologische Intuition bezüglich der *nicht-derivativen Gegebenheit* subjektiver Inhalte. Mentale Präsentate werden so schnell und zuverlässig aktiviert, daß die sie darstellende und zu Bewußtseinsinhalten machende Metarepräsentationsfunktion ihren Konstruktionsprozeß nicht mehr erfaßt.[6] Dadurch werden

[6] Es mag auch sein, daß diese metarepräsentationale Erfassung deswegen nicht gelingt, weil sie bereits in der funktionalen Architektur des Gehirns nicht „vorgesehen" ist.

Präsentate (aber auch andere durch sehr schnelle Operationen erzeugte mentale Zustände) nicht als *Konstrukte*, sondern als *Objekte* bzw. objektive Eigenschaften in das aktuelle Realitätsmodell eingebettet. Dieses repräsentationale Charakteristikum menschlicher Nervensysteme erzeugt eine psychologische Eigenschaft, die man als *naiven Erlebnisrealismus* oder *Gegebenheits-Illusion* bezeichnen könnte. Die entscheidende, sich aus neueren empirischen Erkenntnissen ergebende Tatsache ist, daß die Instantiierung dieser psychologischen Eigenschaft in starker Weise durch bewußtseinsexterne Faktoren determiniert wird. Da diese Eigenschaft durch kontingente physikalische Ereignisse in ihr genaues Gegenteil verkehrt werden kann, verschließt sie sich der philosophischen Interpretation als evidenter *a-priori*-Wahrheit.

Wir sind biologische Wesen, die psychologische Eigenschaften erwerben, indem sie bestimmte mentale Modelle in sich aktivieren. Da mentale Modelle zudem auch die Werkzeuge kognitiver Operationen[7] sind, führen solche Operationen häufig zur Instantiierung neuer psychologischer Eigenschaften. Eine der interessantesten kognitiven Operationen ist die *Selbstbezugnahme*.

> **P 2**: *Welche psychologischen Eigenschaften entstehen dadurch, daß eine Person mental oder extern in Form monologischer Selbstreferenz auf sich Bezug nimmt?*

Um diese Frage zu beantworten, müssen wir mehrere Fälle unterscheiden.

a) *Externe „monologische" Selbstreferenz* ist dann gegeben, wenn ich zum Beispiel *zu mir selbst* sage: „Ich bin gerade etwas verwirrt". Ich produziere in solchen Fällen ein externes Repräsentat, dessen Referenzobjekt das aktuelle *interne* Selbstmodell ist. Deshalb bin ich in solchen Fällen auch gegenüber anderen Sprechern privilegiert, da ich auf größere Informationsressourcen zugreifen kann: In sozialen Situationen kann ich immer nur eine Teilmenge der mir zugänglichen Eigenschaften meiner selbst instrumentalisieren. Durch diesen „monologischen" Sprechakt verändert sich sofort auch der aktuelle Gehalt des von meinem Gehirn aktivierten Selbstmodells: Ich *erlebe* mich nun als einen mit sich selbst Sprechenden, der sich gerade eine psychologische Eigenschaft zuschreibt.

b) *Die mentale Simulation von externer monologischer Selbstreferenz* besteht darin, daß *innerhalb* des Systems eine Serie von mentalen Modellen aktiviert wird, die in analogem Format die Produktion externer sprachlicher Repräsentate nachahmt. Im Normalfall werden diese Modelle auto-

[7] Das intellektuelle Subjekt im klassischen Sinne eines *denkenden Egos* ist das mentale Modell eines mit nicht-sensorischen Repräsentaten höherer Stufen operierenden Selbstes. Es wird durch diejenige Partition des bewußten Selbstmodells gebildet, in die mentale Modelle eingebettet werden, die satzartige Strukturen und propositionalen Gehalt in analogem Format darstellen. Vgl. hierzu auch Rosenberg 1986.

matisch in das Selbstmodell eingebunden. Dadurch entsteht auf der phänomenalen Ebene das subjektive Erlebnis ***„Ich selbst denke gerade,*** *daß ich etwas verwirrt bin*". Propositionale Einstellungen sind der hier vertretenen Auffassung zufolge in das Selbstmodell eingebettete mentale Modelle von propositionalen Repräsentaten. Das subjektive Erlebnis der inneren Produktion von Gedanken im klassischen „assertorischen" Sinn entsteht, wenn die entsprechenden Partitionen des Selbstmodells aktiv sind und durch Metarepräsentation noch einmal abgebildet werden.

c) *Mentale Selbstmodellierung* ist der einfachste Fall – nämlich die im Wachzustand automatisch ablaufende multimodale Selbstrepräsentation des Systems, so wie ich sie im dritten Kapitel beschrieben habe. Hier handelt es sich – ebenso wie im Fall **b)** – *nicht* um Selbstreferenz: Mentale Modelle *referieren* nicht, sie *simulieren*. Die mentalen Modelle des Systems als Ganzem oder der von ihm initiierten Sprechakte *ähneln* ihren Gegenständen, aber sie erzeugen nicht Wahrheit und Falschheit im Sinne wissenschaftlicher Theorien. Die durch Selbstmodellierung erzeugte Perspektivität des inneren Erlebnisraums ist der Kernaspekt derjenigen psychologischen Eigenschaft, die wir üblicherweise als „Subjektivität" zu bezeichnen pflegen. Werfen wir jedoch zunächst noch einen Blick auf die erkenntnistheoretische Struktur interner und externer Selbstrepräsentation.

> **E 2**: *Was ist der erkenntnistheoretische Status des psychologischen Subjekts? Impliziert die Inkorrigibilität von Selbstzuschreibungen psychologischer Eigenschaften ihre Infallibilität?*

Die Eigenschaften des psychologischen Subjekts sind die Eigenschaften einer bestimmten Partition des Selbstmodells. Interne Selbstmodellierung ist aber letztlich eine Aktivität, die man dem System *als Ganzem* zuschreiben muß. Das gilt auch für die Selbstzuschreibung psychologischer Eigenschaften in externen Codes, weil die zugrundeliegenden physikalischen Datenstrukturen vom System als Ganzem produziert werden – und nicht vom psychologischen Subjekt oder vom Selbstmodell (als einem über seinen Gehalt individuierten physischen Teil des Systems). *Infallibilität* im wissenschaftstheoretischen Sinne[8] besitzen Aussagen über die Eigenschaften von Selbstmodellen deshalb nicht, weil Selbstmodelle prinzipiell öffentlich zugängliche Entitäten sind. Das heißt, daß die Analyse eines Selbstmodells – zum Beispiel als einer physikalisch realisierten, aktiven Datenstruktur in einem biologischen Gehirn – prinzipiell von einem Neurowissenschaftler der Zukunft erfolgreicher durchgeführt werden könnte als von dem jeweiligen System selbst, das durch interne Metarepräsentation einen Teil seines aktiven Selbstmodells zu Bewußtseinsinhalten macht. Die Tatsache allein, daß *diese* Form der Darstellung des Selbstmodells eine *interne* ist und das sie von dem fraglichen System *selbst* durchgeführt wird, verleiht

[8] Vgl. Peirce 1893, Albert 1968, Popper 1934b.

diesem noch keine spezielle epistemologische Autorität. Auf der anderen Seite würde uns auch eine komplette neuroinformatische Theorie der Aktivität eines Fledermausgehirns – die wir *mental* niemals erfassen könnten - keinen Zugriff auf das Fledermausbewußtsein ermöglichen, weil die interne funktionale Architektur des Nervensystems der Fledermaus ihr Weisen der kausalen Wechselwirkung mit ihren eigenen Zuständen ermöglicht, die keine theoretische Struktur *uns* liefern kann.[9]

> **E 3**: *Gibt es irreduzible Tatsachen bezüglich der Subjektivität mentaler Zustände, die nur in der ersten Person Singular ausgesprochen und nur aus der Perspektive der ersten Person erkannt werden können?*

Die Eigenschaften subjektiver mentaler Zustände sind die Eigenschaften aktiver mentaler Modelle, die in das Selbstmodell eingebettet sind. Es gibt aber eine wichtige Hinsicht, unter der *nur das die jeweiligen mentalen Modelle aktivierende System selbst* mit Hilfe eines externen, propositionalen Repräsentationscodes auf diese bezugnehmend zugreifen kann: Die Hinsicht der *Internalität*. Wenn ich sage: *„Ich habe gerade Schmerzen“*, dann bin ich das einzige System im Universum, das mit diesem propositionalen Externrepräsentat auf sein aktuelles Selbstmodell Bezug nehmen kann und das sowohl unter der Hinsicht der *Internalität*, als auch unter der Hinsicht der *Identität der die beiden Repräsentate physikalisch*[10] *produzierenden Systeme*. Diese speziellen Eigenheiten der Selbstzuschreibung psychologischer Eigenschaften bringen aber keinen höheren Gewißheitsgrad mit sich – es handelt sich nicht um eine privilegierte Form von *Erkenntnis* aus der Perspektive der ersten Person. Psychologische Selbstzuschreibungen in der ersten Person Singular *(„Ich habe gerade Schmerzen“)* sind Sätze, die durch eine zusätzliche *Hinsicht* gekennzeichnet werden, nicht durch eine eigene Logik.

> **LS 2**: *Was sind die Wahrheitsbedingungen solcher Sätze?*

Sätze, die sich unter der Hinsicht der Internalität auf vom Sprecher intern aktivierte mentale Modelle des Selbst beziehen, werden durch Eigenschaften dieser Modelle wahr oder falsch gemacht. Als neurobiologisch realisierte Datenstrukturen besitzen Selbstmodelle öffentlich zugängliche und

[9] Vgl. Van Gulick 1985: 67. Man kann diesen Gedanken auch über den Begriff der „informationalen Äquivalenz“ zu erläutern versuchen; vgl. a.a.O.: 59.

[10] Ich gehe dabei davon aus, daß die Bedeutung des externen Repräsentats durch einen sozialen Kontext festgelegt wird. Was die Erzeugung von *semantischem Gehalt* angeht, ist der Sprecher also nur Teil eines wesentlich größeren Repräsentationssystems. In bezug auf die physikalischen Datenstrukturen (Schallwellen, Schriftzeichen), die seinem Sprechakt zugrunde liegen, ist er allerdings der alleinige Produzent.

in naturwissenschaftlicher Terminologie beschreibbare Eigenschaften. Als vom System selbst intern noch einmal in analogem Format *metarepräsentierte* Strukturen – also als Inhalte phänomenalen Bewußtseins und repräsentationale Instantiierungsbasis psychologischer Eigenschaften – besitzen sie ebenfalls öffentliche Eigenschaften, die prinzipiell empirisch erfaßt werden können. Trivialerweise führt die Erzeugung solcher Beschreibungen in intersubjektiven Repräsentationssystemen nicht ihrerseits zur Instantiierung psychologischer Eigenschaften durch dieselben.

> **LS 3**: *Läßt die Elimination des Subjektgebrauchs von „Ich" eine Lücke in unserem Verständnis offen?*

Wie Thomas Nagel bemerkt hat[11], befähigt uns die Tatsache, daß wir unpersönliche Wahrheitsbedingungen für Aussagen in der ersten Person angeben können, nicht, diese Aussagen zu machen, ohne die erste Person zu verwenden. Wenn Sie Schmerzen in ihrem Fuß haben und zu mir sagen *„Ich habe Schmerzen in meinem Fuß"*, dann ist das etwas anderes, als zu sagen: *„Dieses Gehirn aktiviert gerade ein Selbstmodell, in das das mit einem Schmerzpräsentat unterlegte mentale Modell eines Fußes eingebettet wurden"*. Was genau ist es, das fehlt? Es ist die eben bereits angesprochene Hinsicht der Internalität: Die Identität von Sprecher und Selbstmodellerzeuger. *Diese* Form der indexikalischen Bezugnahme auf ein mentales Modell kann immer nur von einem einzigen System geleistet werden, in *objektiven* Aussagen über mentale Selbstmodelle muß diese Hinsicht der Internalität zugunsten des von solchen Aussagen angestrebten höheren Allgemeinheitsgrades eliminiert werden. Das, was ich als die Hinsicht der Internalität bezeichnet habe, ist der kleine rote Pfeil auf der Landkarte, der eine objektive Darstellung eines Zustandes der Welt *(„Dieses Gehirn aktiviert gerade ein Selbstmodell, in das das mit einen Schmerzpräsentat unterlegte mentale Modell eines Fußes eingebettet wurde")* in eine nur noch für von einem einzigen Benutzer sinnvoll einzusetzende Repräsentation transformiert *(„Ich habe Schmerzen in meinem Fuß")*. Ist der kleine rote Pfeil ein metaphysisches Mysterium, das auf eine fundamentale und philosophisch bedeutsame Unvollständigkeit des wissenschaftlichen Weltbilds hinweist? Dies ist Thomas Nagels These.

> Es stellt sich damit die Kardinalfrage, ob die Elimination dieses eigentümlichen Gedankens der ersten Person zugunsten seiner impersonalen Wahrheitsbedingungen in unserer Weltbeschreibung nicht doch eine wesentliche Lücke hinterläßt. Und ich bin der Meinung, daß sie auf jeden Fall etwas ausläßt.[12]

Eine wichtige Beobachtung, die sich aus den vorangegangen Überlegungen ergibt, scheint jetzt die folgende zu sein: Es gibt keinen *Subjektgebrauch*

[11] Vgl. Nagel 1992 (1986): 104.
[12] Vgl. Nagel 1992 (1986): 106.

von „Ich“.[13] Was es gibt, ist *der vom System durchgeführte Gebrauch von „Ich“ unter der Hinsicht der Internalität des Selbstmodells.* Mit „Ich“ bezieht sich ein informationsverarbeitendes System bei der Selbstzuschreibung psychologischer Eigenschaften auf ein von ihm intern aktiviertes Analogrepräsentat seiner selbst, das die Basis für die Instantiierung der fraglichen Eigenschaften darstellt. Diese indexikalische Bezugnahme wird natürlich nicht vom phänomenalen Selbst durchgeführt, sondern von dem System als Ganzem. Der Sprecher ist immer das System als Ganzes. Das System als Ganzes ist aber zur gleichen Zeit das Repräsentandum des Selbstmodells, auf das das Wörtchen „Ich“ abzielt. Das, was wir seit Wittgenstein[14] häufig als den Subjektgebrauch von „Ich“ bezeichnen, ist also seiner logischen Struktur nach *die von einem System durchgeführte Indizierung eines Selbstmodells durch die Erzeugung eines externen Repräsentats in propositionalem Format unter der Hinsicht der Internalität*.[15] Diese logisch-epistemische Struktur kann nur von dem betreffenden System selbst realisiert werden, darum ist es in einer prinzipiellen Weise privilegiert gegenüber jeder objektiv-wissenschaftlichen Beschreibung.

Was aber ist die bedeutende Lücke, die durch diese Tatsache in unserer Konzeption der Welt entsteht? Machen wir die starke Annahme, daß die wissenschaftliche Taxonomie mentaler Zustände sich mit der Individuierung mentaler Zustände durch unser Gehirn verknüpfen ließe. Selbst in diesem Fall werden zukünftige wissenschaftliche Aussagen über mentale Selbstmodelle niemals den kleinen roten Pfeil besitzen, der auf der Ebene subjektiven Erlebens die phänomenale Qualität der „Meinigkeit“ erzeugt und auf der Ebene äußeren Sprechens in der Hinsicht der Internalität besteht – das ist auch nicht ihr Ziel. Dieser Punkt wird noch deutlicher werden, wenn wir die folgenden beiden Fragen untersuchen.

LS 6: *Sind Sätze des Typs „Ich bin TM“ ausschließlich als Identitätsaussagen zu interpretieren?*

LS 7: *Gibt es nicht-triviale Wahrheiten in Gestalt solcher Aussagen, die über die reine Selbstreferenz qua historische Person in einem sozialen Kontext hinausgehen? Was wären die Bedingungen der Wahrheit für solche Identifikationen?*

Wie ich im vierten Kapitel dargelegt habe, ist Nagels neuere Reformulierung des Subjektivitätsproblems in ihrem Kern eine These über eine bestimmte Klasse von Sätzen – über Sätze vom Typ „Ich bin TM“. Die

[13] Vgl. Wittgenstein 1970 (1958): 106f, dazu auch Shoemaker 1981b (1968).
[14] Vgl. Wittgenstein 1970 (1958): 106f.
[15] Die dieser Hinsicht zugrundeliegende Internalität ist eine schlicht *räumliche*: Die neurobiologische Realisierung des Selbstmodells befindet sich *im* System (z. B. in seinem Kopf).

offensichtlichen Wahrheitsbedingungen von Sätzen wie „Ich bin TM“ bestehen, wir erinnern uns, in einer Anzahl von Wahrheiten über die historische Person des Sprechers und in einem simplen Kontext der Äußerung: „Ich bin TM“ ist nämlich genau dann wahr, wenn er von TM geäußert wird.[16] Er kann nicht substituiert werden durch eine Dritte-Person-Aussage, aber die seine Wahrheitsbedingungen ausmachenden Fakten können alle ausgedrückt werden durch solche Aussagen. Nagel sagt nun, daß die fraglichen Sätze gewissermaßen einen „doppelten Set von Wahrheitsbedingungen“ besitzen, weil sie eine nicht-triviale Wahrheit ausdrücken, die mit einem irreduziblen und perspektivischen Erste-Person-Faktum zusammenhängt.

Vor dem Hintergrund der im zweiten Kapitel skizzierten Theorie mentaler Repräsentation und der im dritten Kapitel vorgeschlagenen Selbstmodell-Theorie psychologischer Subjektivität kann man nun eine Interpretation der von Nagel vorgeschlagenen „zweiten Lesart“ solcher Sätze anbieten. Wenn solche Sätze in einem *monologisierenden* Kontext geäußert werden[17], dann dienen sie einer *innerpsychischen Orientierung des Subjekts.* Genauer gesagt: Sie indizieren ein spezielles phänomenales Ereignis, das als *repräsentationales* Ereignis beschrieben werden kann. Diese Sätze beziehen sich in solchen Fällen nämlich auf zwei mentale Modelle, von denen das eine in das andere *eingebettet* wird. „Ich“ bezieht sich – in Fällen monologisierender Selbstbezugnahme – auf das aktuelle mentale Selbstmodell der betreffenden Person.[18] „TM“ bezieht sich auf das mentale Modell der objektiven und öffentlichen Eigenschaften (in diesem Fall von Thomas Metzinger). Wie ein Selbstmodell entsteht habe ich im vergangenen Kapitel erläutert. Das mentale Modell von Thomas Metzinger aus der „inneren Außenperspektive“ wird dagegen durch den *View from Nowhere* erzeugt: Indem wir Nagels Anweisungen für einen ganz bestimmten Typ von *mentaler Simulation* befolgen, immer größere Partitionen des Selbstmodells entleeren und in das mentale Modell einer „TM“ genannten Person überführen, geschieht das, was ich im ersten Kapitel als die *Verobjektivierung* aller denkbaren subjektiven Beschreibungen meiner selbst, als das Hinausschieben ihres Gehalts aus dem Ich in eine Totalansicht der Welt *sub specie aeternitatis* paraphrasiert habe. Man darf hier nicht übersehen, daß es sich lediglich um eine mentale Simulation (eine „repräsentationale Verobjektivierung“), also um ein absichtlich durchgeführtes Gedankenexperiment handelt und nicht um die Erzeugung des analogen repräsentationalen *Gesamtzustandes* – dies käme dem Eintritt in ein mystisches Erlebnis gleich.

---

[16] Vgl. Abschnitt 3.2.2

[17] Oder wenn ihre Äußerung durch das Denken eines *philosophical thought* mental simuliert wird.

[18] Das System *kann* sich auf nichts anderes als auf sein Selbstmodell beziehen, weil dies in den zugrundegelegten Situationen die *einzige* Art ist, in der es sich selbst gegeben ist. In monologischen Situationen ist „Ich“ ein Indikator, mit dem ein System sich selbst auf den Gehalt des *jetzt* aktiven Selbstmodells hinweist.

Der *Blick von nirgendwo* ist eine komplexe mentale Operation, die Nagel uns vorschlägt, um uns hinterher eine essentialistische Interpretation der durch sie entstehenden inneren Erfahrungen anzubieten. Diese Operation ist reversibel: Ein System erzeugt **zuerst** in einem Teil seines inneren Simulationsraumes das mentale Modell einer „TM" genannten Person. Dieses mentale Modell reichert es so stark wie möglich mit dem Gehalt eines *anderen* Modells an, nämlich mit dem Gehalt des Selbstmodells (würde das System sein Selbstmodell vollkommen deaktivieren, gäbe es auch kein subjektives Erleben und keine Erinnerung an den *View from Nowhere* mehr). So entsteht das mentale Modell einer anderen Person, allein in Ozeanen von Raum und Zeit, *„a momentary blip on the cosmic TV-screen"*[19] – ein Modell des Systems, als ob es nur durch indirekte Quellen *externen* Wissens gegeben wäre. Man kann diese mentale Operation auch als den Versuch beschreiben, den kleinen roten Pfeil, den ich schon mehrmals erwähnt habe, von einer Simulation[20] der dynamischen, multimodalen Landkarte, die unser Gehirn von der Realität zeichnet, zu entfernen.

In einem **zweiten** Schritt kann das System diesen Pfeil nun wieder einsetzen, beziehungsweise *das mentale Modell der Person „TM" wieder in das Selbstmodell einbetten*. Dieses innere Ereignis ist in der Tat ein erstaunliches psychisches Phänomen, das die Aufmerksamkeit der Philosophie des Geistes – wenn auch keine cartesianisch-husserlsche Interpretation – verdient. Auf *dieses* Ereignis, auf die Fusionierung zweier mentaler Modelle unter sehr speziellen Bedingungen also, beziehen sich Sätze wie „Ich bin TM" in ihrer *zweiten* Lesart. In dieser zweiten Lesart sind sie *monologisierende Identitätsaussagen*, die der innerpsychischen Orientierung des Systems dienen, indem sie indexikalisch auf ein *repräsentationales* Ereignis im System Bezug nehmen. Das System führt eine Integration zweier mentaler Simulate durch, indem es das mentale Modell der Person „TM" wieder in sein Selbstmodell einbettet. Auf diese mentale Operation kann es (in einer monologischen Situation) mit Hilfe des Satzes „Ich bin TM" Bezug nehmen. Die Operation wird auf der Ebene phänomenalen Erlebens von einem Ereignis begleitet, das ich im vierten Kapitel etwas pathetisch als den Absturz aus dem unendlichen Raum, den der *View from Nowhere* in uns öffnet, in die Begrenzungen und Perspektivengebundenheit eines partikularen psychologischen Subjekts beschrieben habe. Allerdings gibt es keinen kleinen Homunkulus, der sich kurzfristig mit dem transzendentalen Ego (Nagels *objektivem Selbst*) vereint hat und nun ins empirische Ich zurückgeschleudert wird. Dies wäre nur eine metaphysische und naiv-realistische Interpretation bestimmter introspektiver Erlebnisse. Die korrekte Interpretation des fraglichen psychischen Phänomens lautet: Ein selbstmo-

[19] Vgl. Nagel 1986: 61.

[20] Man darf dabei nie vergessen, daß es sich nur um intendiertes Imaginieren handelt, also nur um eine *Simulation* des Realitätsmodells. Nach der von mir vorgeschlagenen Theorie mentaler Repräsentation würde eine Entfernung des kleinen roten Pfeils (des Selbstmodells) im *primären* mentalen Realitätsmodell zu einem Zusammenbruch subjektiven Bewußtseins führen.

dellierendes System aktiviert vorübergehend in einem Teilbereich seines inneren Simulationsraums das mentale Modell einer Person, die genau jene Eigenschaften besitzt, die vorher im Modus der *Selbstmodellierung* intern dargestellt wurden. Nach einiger Zeit beendet es diese mentale Simulation und bindet das in ihr entstandene Modell wieder in das Selbstmodell ein. Und auch hier gilt: Es gibt nur ein einziges System, das auf diesen Vorgang *unter der Hinsicht der Internalität* Bezug nehmen kann.

Theoretisch könnte auf die eben angesprochene Weise vielleicht ein reines Ich entstehen, gewissermaßen die pure Selbstmodellierungsfunktion ohne jeden repräsentationalen Gehalt. Aber solange ich mich noch an den *Blick von nirgendwo* erinnern kann, war es auch **mein** *Blick von nirgendwo* – also letztlich nur ein subjektives Erlebnis und keine Berührung der transzendentalen Ebene. Wenn ich vollständig in die außerweltliche Leere eines *objective self* abgeglitten wäre, könnte ich auch nicht mehr den selbstlokalisierenden *philosophical thought* („Ich bin TM") denken[21], der mich zurückholt in die partikulare Perspektive der historischen Person Thomas Metzinger.

> **LS 8**: *Was drückt der „philosophische Gedanke" aus? Was heißt: „Dieses denkende Subjekt betrachtet die Welt durch die Person TN"? Kann man sinnvollerweise davon sprechen, daß das Subjekt die öffentliche Person als Erkenntnisinstrument einsetzt?*

Das, was Thomas Nagel als das „objektive Selbst" bezeichnet, ist die Verdinglichung eines Prozesses, der sich auf der Ebene mentaler Repräsentation ereignet. Dieser Prozeß erzeugt innerhalb eines perspektivischen Repräsentationsraums (innerhalb des *subjektiven* Bewußtseins einer Person, die mit dem *Blick von nirgendwo* experimentiert) ein *aperspektivisches*

[21] Daß wir aus Zuständen ohne die Instantiierung psychologischer Eigenschaften (zum Beispiel Tiefschlaf oder Koma) als phänomenale Subjekte wieder auftauchen können, ist ein starkes Argument gegen strikt *intramentale* Ursachenketten, wenn man für sie das „Prinzip der zeitlichen Nachbarschaft" von Ursache und Wirkung voraussetzt. Die Episodizität unseres Mentallebens deutet auf seine Supervenienz gegenüber dem Physikalischen hin. Das Problem der *intramentalen* Kausalität war interessanterweise sogar schon für die frühbuddhistische Philosophie des Geistes von zentraler Bedeutung, denn es gab eine ausgedehnte Diskussion zwischen den philosophischen Traditionen des Theravada, des Vaibhasika und des Yogacara. Thema dieser Debatte war ein Bewußtseinzustand, der von großer soteriologischer Bedeutung war und den wir aus der Perspektive westlicher Taxonomien vielleicht als „kataleptische Trance" bezeichnen würden: Das „Erreichen des Erlöschens" *(„the attainment of cessation")*. Es ist ein Zustand, in dem keinerlei mentale Ereignisse stattfinden. Personen, die sich in diesem Zustand befinden, haben *prima facie* keine psychologischen Eigenschaften. Das wirft natürlich die Frage auf, inwieweit sie überhaupt noch Personen sind bzw. – aus heutiger Perspektive – was es genau heißt, daß ein informationsverarbeitendes System seine Psychologie vorübergehend völlig verliert. Eine gute Darstellung der frühbuddhistischen Debatte, die sich bereits auf einem recht hohen analytischen Niveau bewegt, und ihrer Beziehung zum Leib-Seele-Problem gibt Griffiths 1986. Vgl. auch Oetke 1988.

mentales Simulat der Wirklichkeit. Dieses nicht-zentrierte mentale Modell der Realität enthält, so sagt Nagel, alle Erfahrungen und die Perspektive von TN:

> Doch die Erlebnisse und die Perspektive von TN, die mir unmittelbar gegeben sind, machen nicht den Standpunkt des eigentlichen Selbst aus, da das eigentliche Selbst eben keine Perspektive hat, sondern TN und seine Perspektive in seiner Auffassung von einer zentrumlosen Welt als Inhalt dieser Welt miteinschließt.[22]

Aber das ist falsch: *Diese* innere Erfahrung, nämlich den aus der Perspektive des psychologischen Subjekts TN initiierten und durchgeführten *View from Nowhere*, enthält die *„centerless conception of the world"* eben nicht.[23] Die *aktuelle* Perspektive ist in dem zentrumlosen Bild der Welt, auf welches das von Nagel postulierte *wahre* Selbst in nicht-perspektivischer Weise schaut, nicht enthalten. Das ist deshalb so, weil der eigentliche Agent, der alle hier diskutierten phänomenalen Zustände und auch ihre philosophischen Interpretationen erzeugt, das Gehirn eines biologischen Organismus ist.

Vielleicht verstehen wir besser, was Nagel uns sagen will, wenn wir eine der Metaphern untersuchen, die er uns anbietet: *„. . . this thinking subject regards the world through the person TN"*.[24] Die wirkliche Situation ist natürlich eine andere: Ein biologisches Repräsentationssystem „betrachtet" die Welt, indem es ein internes Modell von ihr und von sich selbst *in* ihr erzeugt. Dieses System bemerkt nun an sich eine interessante Fähigkeit: Es kann sich von seinem mentalen Selbstmodell immer wieder distanzieren, indem es Teile davon mental als *Objekt* beziehungsweise als eine *andere* Person modelliert. Als natürlicher Psychologe versucht dieses System aber auch immer, sich seine eigenen kognitiven Operationen und Kapazitäten zu erklären. Dies tut es jedoch unter ganz bestimmten Bedingungen: Unter den Bedingungen eines naiv-realistischen Selbstmißverständnisses, welches durch die semantische Transparenz der beteiligten mentalen Modelle erzeugt wird. Das führt dazu, daß es sich als ein Wesen begreift, das sich von seinem phänomenalem Ich (dem Selbstmodell) immer wieder distanzieren kann, indem es dieses mit Hilfe des *Blicks aus dem Nirgendwo* in ein „objektives" inneres Bild hinüberschiebt. Daraus schließt das System korrekterweise, daß es nicht mit diesem phänomenalen Ich identisch ist, sondern es nur als Erkenntnisinstrument *benutzt*. Die naiv-realistische Interpretation dieses Sachverhalts lautet: *„Ich betrachte die Welt durch die historische Person TN"*. Dann aber entsteht das folgende Problem: Wer bin *ich*? Wer ist es, der eben *diese* Operation durchführt? Das, was vor dem

[22] Vgl. Nagel 1992: 109.

[23] Wenn es sie doch in ihrem vollen Gehalt enthielte, würde eine Erfahrung entstehen, die TN *im Nachhinein* folgendermaßen beschreiben könnte: *„Ich habe mich vollkommen aufgelöst und bin mit dem objektiven Selbst verschmolzen"*. Das ist aber nicht Thomas Nagels These.

[24] Vgl. Nagel 1986: 60.

Hintergrund einer naturalistischen Theorie der mentalen Repräsentation eben *das physikalische System als Ganzes* ist, wird nun in der reifizierenden Interpretation dieses mentalen Repräsentationsprozesses zum *thinking subject*: In Ermangelung einer befriedigenden Erklärung für diese immer schon vorhandene Fähigkeit zur repräsentationalen Übersteigung des Selbstmodells gelangen wir somit fast zwangsläufig zur Postulierung eines transzendentalen Subjekts.

Genauso irreführend wie diese essentialistische Metaphorik wäre auch die These, daß das Subjekt die öffentliche Person als Erkenntnisinstrument einsetzt. Denn es ist das *mentale Modell* der Person, welches für diese zum Erkenntnisinstrument wird. Das Subjekt – der Produzent von *cogitationes* und von Selbsterkenntnis – ist ein biologisches System, das gewisse abstrakte Organe gebildet hat und diese nun in sich aktiviert. Wir werden diesen Punkt noch besser verstehen, wenn wir einige der Fragen beantworten, mit denen wir im ersten Kapitel die *psychologische* Dimension des Subjektivitätsproblems definiert haben.

> **P 3**: *Was ist der phänomenale Gehalt mentaler Zustände im Gegensatz zu ihrem intentionalen Gehalt? Gibt es mentale Zustände ohne phänomenalen Gehalt?*

Allgemein gesprochen entsteht phänomenaler Gehalt dadurch, daß ein System mentale Repräsentate, Simulate und Präsentate metarepräsentational noch einmal abbildet und damit zu *bewußten* Zuständen macht: Phänomenale Zustände sind Bewußtseinszustände. Über die neurobiologischen Mechanismen, die den fraglichen Operationen in unserem eigenen Fall zugrunde liegen, kann man derzeit noch wenig aussagen. Die Netzwerktheorie hat uns jedoch gezeigt, daß Metarepräsentation auch *subsymbolische* Metarepräsentation sein könnte, die nicht regelgeleitet und nur schwach kompositional[25] ist – also *subkognitiv* im Sinne klassisch-kognitivistischer Theorien à la Fodor und Pylyshyn.

In dem speziellen Sinn mentaler Präsentation (wie ich ihn in Abschnitt 2.1.3 entwickelt habe) bedeutet das Auftreten von phänomenalem Gehalt die Aktivierung von *Präsentaten* durch das System. Mentale Präsentate sind nicht-relationale Datenstrukturen, die Signalcharakter haben und aufgrund ihres Formats und ihrer Funktion *für* das System als Analog-Indikatoren beschrieben werden können. Der spezielle qualitative Gehalt der durch Analog-Indikation im System erzeugten Datenstrukturen wird durch ihr *Format* hervorgebracht: Phänomenale Klänge oder phänomenale Farben sind abstrakte Eigenschaften der physikalischen Realisierung dieser Datenstrukturen, die vom Gehirn erkannt werden.

[25] Vgl. Goschke/Koppelberg 1990.

P 4: *Koinzidiert das Auftreten von phänomenalem Gehalt immer mit dem, was wir „Bewußtsein" nennen? Gibt es Gegenbeispiele?*

Es gibt Gegenbeispiele: Das *unbewußte Farbensehen* von Blindsicht-Patienten etwa, die innerhalb ihres Skotoms präsentierte Farben unterscheiden können. Es gibt also erfolgreiches diskriminatives Verhalten bezüglich Farbunterscheidungsaufgaben *ohne* bewußtes Seherlebnis.[26] Im Rahmen der hier skizzierten Theorie könnte man vermuten: In solchen Fällen werden funktional aktive mentale Präsentate erzeugt, die vom System aufgrund der vorliegenden organischen Schädigung nicht mehr metarepräsentational abgebildet werden können.

E 5: *Stellt der qualitative Gehalt mentaler Zustände eine Form von „Wissen über die Welt" dar? Sind Qualia Informationsträger, gibt es so etwas wie „phänomenale Information"?*

Mentale Präsentate sind Informationsträger. In den Standardsituationen informieren sie das System auf sehr schnelle Weise über die *pure Präsenz* einer Reizquelle. Sie produzieren eine sehr einfache Form von Wissen, wenn man so will, unterstützen sie „Existenzannahmen" innerhalb eines Realitätsmodells, indem sie bestimmte mentale Modelle als *über eine spezifische Signalquelle gegeben* auszeichnen. Qualia sind Informationsträger, aber es ist – wie das eben angeführte Beispiel uns zeigt – höchstwahrscheinlich nicht die Tatsache, daß sie Inhalte phänomenalen Bewußtseins sind, die ihnen ihre funktionale Rolle innerhalb des Systems verleiht. Die Information, die sie für das System intern präsentieren, ist also keine *besondere* Art von Information – sie ist nicht essentiell an die Koinstantiierung mit phänomenalen Eigenschaften gebunden. In Nicht-Standardsituationen (Träumen, Halluzinationen) kann phänomenaler Gehalt auch durch interne Signalquellen erzeugt werden und dann zum Entstehen von psychischen Artefakten führen. Eine Person, die die korrekte wissenschaftliche Theorie über mentale Präsentate kennt, könnte sich über die kausale Genese der von ihr gesehenen Farben täuschen – nicht aber über die Art der jeweils aktuellen Systemzustände: Wenn Frank Jacksons Mary ein Roterlebnis hat, mag sie halluzinieren, aber sie weiß genau, daß ihr visueller Cortex momentan ganz bestimmte Mikrozustände erzeugt, die (als neurobiologisch realisierte Datenstrukturen betrachtet) in ganz spezifischen Formaten vorliegen. In *diesem* Sinne präsentieren Qualia dem System in jedem Fall Information über seine eigene innere Struktur.

[26] Vgl. Cowey/Stoerig 1991b, Pöppel 1987, Stoerig/Cowey 1990, 1991a, 1992, 1993, Weiskrantz 1986.

LS 4: *Kann die Individuierung mentaler Zustände über intrinsische, nicht-relationale Eigenschaften wie ihren qualitativen Gehalt im Rahmen einer wissenschaftlichen Psychologie aufgegeben werden?*

Die „intrinsischen" und „nicht-relationalen" Eigenschaften, anhand derer das Gehirn einzelne seiner präsentationalen Zustände herausgreift, sind aller Wahrscheinlichkeit nach variabel. Wenn die zugrundeliegende Informationsverarbeitung massiv-parallel ist, dann kann man auch davon ausgehen, daß die durch sie erzeugten Aktivierungszustände und Datenstrukturen niemals genau dieselben, also keine „repräsentationalen Atome" sind. Dafür spricht auch die phänomenale *Plastizität* von Qualia: Unser Vermögen, Gerüche, Klänge oder Farben zu unterscheiden, variiert inter- und intraindividuell stark – man kann es sogar in nicht unbeträchtlichem Umfang trainieren. Das heißt: Die entsprechenden Formate sind ebenfalls plastisch. Es ist wohl eher unwahrscheinlich, daß die theoretische Taxonomie einer möglichen zukünftigen wissenschaftlichen Beschreibung mentaler Präsentate sich ohne weiteres auf die (unscharfe) metarepräsentationale Taxonomie dieser Zustände abbilden läßt, die das Gehirn erzeugt und auf die wir dann in externen Codes als „Qualia" Bezug nehmen.

Deshalb muß die traditionelle Weise der Individuierung mancher mentaler Zustände über ihren qualitativen Gehalt früher oder später zugunsten anderer Merkmale aufgegeben werden. Wenn dies geschieht, wird es sich letztlich auch in unserer Alltagspsychologie niederschlagen. Zur Entscheidung über die „erkenntnistheoretische Autorität" des introspizierenden Subjekts und der mit ihm konkurrierenden wissenschaftlichen Theorien wird man den Erfolg bei der Voraussage kausaler Konsequenzen des Auftretens bestimmter mentaler Ereignisse heranziehen: Wenn die Wissenschaft das funktionale Profil der fraglichen inneren Zustände signifikant besser beschreiben kann, dann wird sich ganz von selbst schließlich auch unsere lebensweltliche Taxonomie dieser Zustände ändern – wir werden auch im Alltag *anders* über subjektive Erlebnisqualitäten sprechen.

Sogar wenn man der *folk psychology* den Status einer Theorie zubilligt, bleibt allerdings fraglich, ob es sinnvoll ist, hier überhaupt von einem intertheoretischen Konkurrenzverhältnis zu sprechen. Individuelle Gehirne und Wissenschaftlergemeinschaften sind sehr verschiedene Repräsentationssysteme, die Information auf sehr unterschiedliche Weise und in der Verfolgung sehr unterschiedlicher Interessen verarbeiten. Gerade weil die Interessen beider Systemtypen divergieren, ist es auch fragwürdig, die eben angesprochene Abbildung von einem der beiden Systeme im Sinne einer epistemischen Rechtfertigung zu fordern.

E 8: *Können introspektive Berichte über innere Eigenschaften in Konkurrenz treten mit Aussagen einer wissenschaftlichen Psychologie?*

Was sind introspektive Berichte über innere Eigenschaften? Es sind sprachliche Äußerungen, mit denen eine Person sich auf *bewußte* mentale Modelle bezieht. In einem weiteren Sinne gehören dazu auch alle bewußten mentalen Modelle, die in das Realitätsmodell der Person eingebettet sind: Die Eigenschaften des Realitätsmodells sind in einem repräsentationalen Sinn die Eigenschaften *interner* Datenstrukturen, also innere Eigenschaften. Die philosophisch interessante Form von *subjektiver Innerlichkeit* wird aber nur durch solche Strukturen erzeugt, die in das aktuelle Selbstmodell eingebettet sind. Denn nur die in das aktuelle Selbstmodell eingebundenen mentalen Modelle instantiieren innere Eigenschaften im Sinne *phänomenaler* Innerlichkeit. Introspektive Berichte und Selbstzuschreibungen psychologischer Eigenschaften beziehen sich also auf diejenige Partition des Selbstmodells eines Sprechers, die durch Metarepräsentation zum Inhalt innerer Bewußtheit geworden ist.[27] Allgemein kann man sagen: Die Aussagen, die ein System introspektiv über seine aktuellen psychologischen Eigenschaften macht, sind indexikalische Aussagen über eine ganz bestimmte Teilmenge der gegenwärtig von ihm aktivierten mentalen Modelle *unter der Hinsicht der Internalität*.

Bei wissenschaftlichen Sätzen über die mentalen Modelle eines Systems oder einer bestimmten Klasse von Systemen fällt die zuletzt genannte Hinsicht weg: Sie können prinzipiell von beliebigen Sprechern geäußert werden. Das liegt daran, daß die von einer wissenschaftlichen Psychologie entworfenen Landkarten der repräsentationalen Struktur, die biologische (und vielleicht in der Zukunft auch einmal artifizielle) Systeme in sich entwickeln, den „kleinen roten Pfeil“ nicht besitzen. Das heißt aber nicht, daß sie uns nicht beschreiben und erklären können, wie die kleinen roten Pfeile in den mentalen Landkarten – von denen die *wissenschaftlichen* Landkarten handeln – entstehen, welche Funktion sie für das System besitzen und welche neuen psychologischen Eigenschaften durch diese „repräsentationale Benutzerfixierung“ ins Spiel kommen. Die wissenschaftliche Psychologie hat als eine empirisch fundierte Theorie natürlicher Repräsentationssysteme außerdem ein weiteres Spektrum als die Alltagspsychologie. Denn sie hat mit ihrem Instrumentarium auch einen direkten Zugriff auf solche mentalen Modelle, die Versuchspersonen introspektiv nicht zugänglich sind – auf das, was man früher „das Unbewußte“ nannte.

Die Alltagspsychologie[28] der ersten Person benutzt ein Erkenntnisinstru-

[27] Aussagen über dispositionale psychologische Eigenschaften („Ich habe einen Hang zur Selbstverliebtheit“) sind dementsprechend Aussagen über die funktionale Struktur der Selbstmodellierung und nicht über den aktuellen Gehalt des durch sie aktivierten Selbstmodells.

[28] Mir fehlt hier der Raum, um auf die Debatte um den Theoriestatus der vorwissenschaftlichen *folk psychology* einzugehen. Die prominentesten philosophischen Gegner der Alltagspsychologie sind Patricia und Paul Churchland, vgl. Churchland, P.M. 1979, 1989, Churchland, P.S. 1986. Verteidigungen des *common-sense*–Intentionalismus finden sich bei Kitcher 1984, von Eckardt 1984, wichtige weitere Texte sind Dennett 1987 und Stich 1978 (1990), 1983. Eine knappe Einführung mit weiteren bibliographischen Angaben und einigen neueren Texten findet sich bei Lycan 1990, Teil VI.

ment, das sie selbst nicht kennt.[29] Es sind die Metarepräsentationsvorgänge des Gehirns, die wir zu steuern meinen, wenn wir die Aufmerksamkeit unseren subjektiven Zuständen zuwenden. In solchen psychischen Situationen aktiviert das Gehirn häufig ein Modell des Selbst als eines inneren (aber höchstens diffus lokalisierten) Homunkulus, der den Lichtstrahl fokussierter Aufmerksamkeit durch die phänomenale Welt wandern läßt. Es mag wahr sein, daß solche Formen der Selbstmodellierung letztlich bloß phänomenale Konfabulationen sind, mit denen das System sich seine eigenen höherstufigen Zustände zu erklären versucht. In unserem Zusammenhang ist nur wichtig, daß Introspektion ein schlecht verstandener Vorgang ist, der eine bestimmte Klasse von mentalen Modellen erzeugt. Mentale Modelle als Analogrepäsentate besitzen keine Wahrheitswerte, und können schon deshalb nicht in Konkurrenz treten mit wissenschaftlichen Theorien – gleich welcher Art. Mentale Modelle erzeugen nicht-diskursives Wissen. Die *Sätze*, mit denen sich eine Person auf die gegenwärtig in ihr aktivierten mentalen Modelle bezieht, werden normalerweise in anderen Kontexten geäußert und verfolgen ein anderes Ziel als die Sätze einer wissenschaftlichen Psychologie.

Wenn Sie sagen: *„Ich spüre, wie sehr ich meinen Mann liebe, weil ich gerade eifersüchtig bin“*, dann könnte ein Neurowissenschaftler der Zukunft Ihnen vielleicht entgegnen: *„Tut mir leid, aber meine Instrumente zeigen mir, daß die von Ihrem Gehirn derzeit durchgeführte emotionale Selbstmodellierung nach der von der Weltgesundheitsorganisation zugrundegelegten wissenschaftlichen Taxonomie mentaler Zustände eindeutig in die diagnostische Kategorie „neurotische Verlustangst“ fällt. Mit Liebe hat ihr innerer Zustand nichts zu tun.“* Wie würden Sie sich in einer solchen Situation verhalten? Wenn es stimmt, daß mentale Modelle funktional aktive Strukturen innerhalb eines Systems sind, die als solche einen Einfluß auf sein inneres und äußeres Verhalten besitzen, dann lassen sich solche Streitigkeiten zwischen introspektiven und wissenschaftlichen Theorien über mentale Zustände dadurch schlichten, daß man ihre *predictive power* miteinander vergleicht, ihre Kraft, erfolgreiche Voraussagen über das zukünftige Verhalten des Systems zu machen. Die erkenntnistheoretische Autorität eines Systems bezüglich seiner eigenen subjektiven Zustände kann nur dadurch zustande kommen, daß es bessere Voraussagen über die kausalen Konsequenzen dieser Zustände machen kann als *andere* Repräsentationssysteme. Ich möchte diesen Punkt jedoch nicht weiter vertiefen, sondern noch einmal auf den psychologischen Aspekt des Problems subjektzentrierter, phänomenaler Zustände zurückkommen. Werfen wir dazu erneut einen Blick auf die drei phänomenologischen Säulen von Subjektivität, so wie ich sie im ersten Kapitel eingeführt habe.

[29] Vgl. hierzu auch den Begriff des *metapsychologischen Wissens* bei Robert van Gulick in Van Gulick 1988a, 1988b.

P 5: *Ist das Auftreten von Qualia an eine subjektive Erlebnisperspektive oder an bestimmte Klassen von Bewußtseinszuständen gebunden? Kann es nicht-subjektzentrierte Qualia geben?*

Kann es Systeme geben, die mentale Präsentate aktivieren, ohne gleichzeitig ein Selbstmodell zu erzeugen? Nach der im zweiten Kapitel skizzierten Theorie ist es durchaus möglich, daß es Systeme gibt, die keine Selbstmodellgeneratoren sind und trotzdem ein großes Repertoire von funktionalen Zuständen besitzen, die die Rolle von Analog-Indikatoren spielen (vielleicht sind eine ganze Reihe von einfachen Biosystemen auf unserem Planeten dieser Kategorie zuzuordnen). In einem solchen System könnte es phänomenale Qualitäten vom Typ „Schmerzen" geben, ohne daß die durch sie begleiteten Zustände *subjektive* Schmerzen wären: Solche Zustände wären *niemandes* Schmerzen. Repräsentational sind solche Zustände auf jeden Fall denkbar. Ebenso wie die unbewußten Farbwahrnehmungen von Blindsichtpatienten, die ich bereits mehrfach erwähnt habe, kollidiert die Möglichkeit der Existenz solcher Zustände frontal mit den Intuitionen der meisten Menschen. Das liegt daran, daß wir uns Schmerzen oder Lusterlebnisse, die niemandes Schmerzen oder Lusterlebnisse sind, nicht *vorstellen* können: Wir können solche Zustände nicht mental simulieren, weil *unsere* Form phänomenalen Bewußtseins automatisch mit der Aktivierung eines sehr komplexen und stabilen Selbstmodells einhergeht. Für das Auftreten bewußter Qualia sind aber nur Metarepräsentation und Präsentation erforderlich. Aus diesem Grund sind nicht-subjektzentrierte Qualia in einem psychologischen Sinn für uns undenkbar, wir sind Wesen, die solche einfachen Formen des Bewußtseins nicht mehr absichtlich in sich erzeugen können. Natürlich kann man auf begrifflicher Ebene anthropozentrisch dafür argumentieren, daß nicht an ein phänomenales Selbst gebundene mentale Zustände Subjektivität nur in einem deprivierten und beschränkten Sinne instantiieren. Auf Saul Kripkes Argument gegen die Identitätstheorie werde ich weiter unten noch einmal zurückkommen.

P 6: *Kann es Bewußtsein ohne qualitativen Gehalt geben?*

Auch hier fällt es schwer, sich entsprechende Zustände vorzustellen: Bewußtseinszustände, die zwar *meine* sind, aber keine Form phänomenalen Gehalts besitzen. Man könnte an Mathematiker im Mathematikerhimmel denken, die ein subjektives Bewußtsein besitzen, dessen Inhalte ausschließlich *Abstrakta* sind – Formen mentalen Gehalts, die die eigentümliche Konkretheit nicht besitzen, die in *unserer* Bewußtseinsform durch sensorische und emotionale Qualitäten erzeugt wird. Ganz allgemein wird man sagen müssen: Qualiafreie Bewußtseinszustände könnten von Systemen generiert werden, die Sinneswahrnehmungen und die interne Modellierung ihrer Interessenlage durch ganz andere funktionale Zustände organisiert

haben als wir. Man kann sich Systeme denken, die zwar Selbstmodelle und reichhaltige metarepräsentationale Zustände in sich aktivieren, aber keine Analog-Indikatoren besitzen. Die mentalen Realitätsmodelle solcher Systeme würden aus relationalen Strukturen ohne den puren „Signalaspekt" mentaler Präsentation bestehen. Repräsentational *möglich* sind solche Systeme – ohne Emotionen und konkrete Sinnlichkeit – auf jeden Fall.

Woran liegt es, daß wir psychologische Subjekte dieser Art als hochgradig „unmenschlich" empfinden würden? Es mag unter anderem daran liegen, daß wir die Biologizität und das spezifische Charakteristikum unserer phänomenalen Realität intuitiv in der *Konkretheit* lokalisieren, mit der uns unsere eigenen Interessen als Gefühle, unsere Sinneswahrnehmungen als Farben, Gerüche oder Klänge, und wir uns *selbst* als erlebnismäßig sehr direkt und unhintergehbar *verkörperte* Wesen[30] gegeben sind. Und es ist ja auch eben diese phänomenale Konkretheit, die es so schwer macht, uns selbst als informationsverarbeitende Systeme zu betrachten. Eine Welt ohne mentale Präsentate wäre eine Welt ohne *Präsenz*: Eine Welt ohne Gegenwart, ohne den Widerstand der Objekthaftigkeit, eine Welt, in der sogar wir selbst nicht in vollem Sinne *anwesend* wären.

P 7: *Kann es aperspektivisches Bewußtsein geben?*

Wir kehren nun zurück zum thematischen Kern der Diskussion, zur *Perspektivität* phänomenalen Bewußtseins. Könnte es ein Bewußtsein ohne Perspektive geben, ein Bewußtsein, bezüglich dessen sich Thomas Nagels *perspectival-facts*-These nicht aufstellen ließe? Nach der hier vertretenen Theorie sind solche Zustände möglich, nämlich dann, wenn ein System ein Realitätsmodell konstruiert, in das kein Selbstmodell eingebettet ist, welches aber ganz oder teilweise durch eine einheitliche Metarepräsentationsfunktion intern abgebildet wird. Solche nicht-zentrierten Realitätsmodelle können durchaus einer teleofunktionalistischen Analyse zugänglich sein, das heißt eine Funktion *für* das System erfüllen. Es gibt aber in ihnen keinen „kleinen roten Pfeil" und auch nicht sein phänomenales Gegenstück, das psychologische Subjekt: Die Inhalte eines aperspektivischen Bewußtseins sind keine *subjektiven* Bewußtseinsinhalte. Ihnen fehlt die repräsentationale und phänomenale Einbindung in einen inneren Brennpunkt, der die psychische Identität des Systems fixiert. Wenn mentale Selbstmodellierung und die Erzeugung einer phänomenalen Person (vgl. Abschnitt 3.2.1) eines unserer Kriterien dafür ist, ob wir einem System den Status des Personseins zubilligen, dann sind Großrechner, viele Tiere, alle Säuglinge und manche Heilige keine Personen. Solche Systeme können sich keine psychologischen Eigenschaften unter der Hinsicht der Internalität des Selbstmodells zuschreiben, und sich auch nicht über den „philosophischen Geschmack" von Sätzen des Typs *„Ich bin TM"* wundern.

[30] Ich erinnere in diesem Zusammenhang an das von mir in Abschnitt 3.2.2 vorgestellte Beispiel der körperlosen Frau.

P 8: *Wie entsteht ein zentriertes Bewußtsein? Wie ist in bestimmten informationsverarbeitenden Systemen die Emergenz eines phänomenalen Standpunkts bzw. einer Erlebnisperspektive möglich?*

Nach der Selbstmodell-Theorie der Subjektivität liegt die entscheidende repräsentationale Instantiierungsbasis für die entsprechenden psychologischen Eigenschaften in der Tatsache, daß ein System beginnt, sich selbst mental zu modellieren und die durch diesen Prozeß entstehende Datenstruktur in sein inneres Bild der Wirklichkeit einzubetten. Der vom System erzeugte mentale Raum erfährt dadurch eine *Zentrierung*: Da das diesen Raum erzeugende und ihn benutzende System sich nun selbst innerhalb dieses Raums in Gestalt eines privilegierten mentalen Modells gegeben ist, verändert sich dessen Gesamtstruktur. Er ist nun nicht mehr nur ein internes Realitätsmodell, sondern sogar ein *benutzerfixiertes* Realitätsmodell. Wenn durch einen teuflischen Neurowissenschaftler der Zukunft die zentrierten Realitätsmodelle zwischen den Gehirnen von mir und einem meiner Leser ausgetauscht würden, verlören wir beide nicht nur unsere psychische und neurobiologische Identität. Wir wären jetzt auch Systeme, deren aktuelle Realitätsmodelle hochgradig *afunktional* sind, weil der „kleine rote Pfeil" in ihnen – also das die Benutzerfixierung leistende Selbstmodell – zwar *ex hypothesi* noch vorhanden, aber seinem Gehalt nach schlicht leer wäre.

Durch die Zentrierung mit Hilfe eines Selbstmodells können interne Simulationsräume aber auch zur Grundlage einer veränderten Psychologie des Systems werden: Selbstmodellerzeuger instantiieren neue psychologische Eigenschaften.

P 9: *Wodurch werden manche Resultate interner Informationsverarbeitung zu* ***meinen*** *Zuständen? Wie entsteht die „Meinigkeit" mentaler Zustände?*

Diejenigen Sequenzen von internen Zuständen eines Systems, die als Informationsverarbeitungsprozesse beschrieben werden können, sind *seine* nur im trivialen Sinne physikalischer Internalität. Manche Systeme – Selbstmodellgeneratoren, die natürliche Sprachen sprechen – nehmen dagegen auf einige ihrer inneren Zustände in externen Codes unter einer anderen Hinsicht bezug: unter der Hinsicht phänomenaler Innerlichkeit. Die Zustände, auf die sie sich beziehen, wenn sie sagen *„Ich glaube an die Unsterblichkeit der Seele"*[31] oder *„Ich bin gerade etwas verwirrt"* sind Zustände des Selbst-

[31] Man kann propositionale Einstellungen nun auch als Zustände des Selbstmodells analysieren, die im Zusammenhang von mentalen Simulationen möglicher Sprechakte entstehen. Das bedeutet: Man analysiert nicht mehr die Beziehung zwischen Person und Proposition, sondern diejenige zwischen System und Selbstmodell.

modells, Zustände eines mentalen Makro-Repräsentats, in das verschiedene andere Datenstrukturen eingebettet wurden. Der Umstand, daß diese Datenstrukturen in das Selbstmodell eingebettet wurden, verleiht ihnen ein zusätzliches Merkmal. Wird dieses Merkmal durch das Gehirn metarepräsentational noch einmal abgebildet, dann entsteht auf der Erlebnisebene die Qualität der „Meinigkeit". Alle mentalen Zustände, die diese phänomenale „Meinigkeit" aufweisen, erzeugen zusammen etwas, das man als das phänomenale Selbst, das Ichgefühl oder als die *subjektive Identität* bezeichnen könnte: Eine durch eine gemeinsame phänomenale Eigenschaft gebündelte Klasse innerer Zustände.

„Eingebettetsein-in-ein-Selbstmodell" ist eine relationale Eigenschaft mentaler Modelle, deshalb habe ich den durch diese Eigenschaft erzeugten phänomenalen Aspekt von Subjektivität im ersten Kapitel auch als relationale Subjektivität bezeichnet. Einbettungsrelationen zwischen mentalen Modellen und die durch sie instantiierten psychologischen Eigenschaften sind ein weites Feld für zukünftige Analysen.[32] Zum Beispiel ist auch das Selbstmodell in eine höherstufige repräsentationale Struktur integriert, nämlich in das *Realitätsmodell* des Systems – vielleicht könnte man die phänomenale Qualität, die dieser Umstand unserer inneren Seinsverfassung in Standardsituationen verleiht, als *In-der-Welt-Sein* bezeichnen. Auch mentale Präsentate werden oft in mentale Modelle eingebunden und verleihen diesen durch den Signalaspekt die Qualität der Konkretheit und Präsenz, die mentale Simulate meist nicht besitzen. Es ließen sich viele Beispiele für mit spezifischen repräsentationalen Einbettungsrelationen korrelierende phänomenale Eigenschaften finden. Dies jedoch ist hier nicht mein Ziel.

„Eingebettetsein-in-ein-Selbstmodell" ist eine Eigenschaft, die in pathologischen Situationen verlorengehen oder den falschen mentalen Modellen vom System verliehen werden kann. In Abschnitt 3.2.2 habe ich einige solcher Fälle diskutiert, um die Tatsache zu illustrieren, wie stark die Instantiierung der fraglichen Eigenschaften auf der Ebene des Selbsterlebens durch Faktoren auf tieferliegenden Beschreibungsebenen determiniert wird. Wenn alle repräsentationalen Eigenschaften eines Systems feststehen,

[32] Im Personen-Kapitel von Strawsons „Individuals" wird die *logische* Beziehung zwischen dem Subjekt und den ihm zugeschriebenen Bewußtseinszuständen (die ich hier mit einer *repräsentationalen* Einbettungsbeziehung zwischen mentalen Modellen zu verknüpfen suche) auch als *Besitzbeziehung* diskutiert: Was ist der korrekte Sinn, indem man sagen kann, daß ein Erlebnissubjekt seine Bewußtseinszustände *besitzt*? Vgl. Strawson 1972 (1958): 111ff. Interessant ist in diesem Zusammenhang vielleicht eine phänomenologische Beobachtung, die man bei manchen Psychosen machen kann – nämlich solchen geistigen Erkrankungen, die früher und leider in seltenen Fällen auch heute noch als Fälle von „Besessenheit" interpretiert wurden: Ein konstruiertes mentales Modell einer fiktiven Person versklavt das ursprüngliche Selbstmodell und führt auf der phänomenalen Ebene eine *Umkehrung* der Besitzrelation herbei. Auch in den gegen Ende des dritten Kapitels geschilderten *Multiple Personality Disorders* (vgl. Abschnitt 3.2.2) scheint es eine „Konkurrenz" oder einen „Wettlauf" um die Realisierung der fraglichen Besitz- oder Einbettungsbeziehung zu geben.

sind auch der Gehalt seines Selbstmodells und damit die Eigenschaften des psychologischen Subjekts fixiert. Die meisten von uns haben noch keine ernsthaften Störungen mentaler Selbstmodellierung erlebt, deshalb können wir uns auch nur schlecht vorstellen, daß unser Ichgefühl im Grunde eine sehr zerbrechliche psychische Instanz ist, die in ihrer Kontinuität von bewußtseinsexternen Faktoren abhängig ist. Es fällt in diesem Zusammenhang auf, daß mentale Simulationen, in deren Verlauf das Selbstmodell deaktiviert wird, praktisch kaum durchführbar sind: Wir können uns nicht vorstellen, wie es ist, wenn wir tot sind. Die Erklärung für diese Tatsache mag sein, daß solche mentalen Operationen biologisch sinnlos oder sogar gefährlich sind.

> **P 10**: *Koinzidiert die fragliche Relation mit dem Auftreten von „Bewußtsein"; kann es unbewußte „Meinigkeit" geben?*

Beim Menschen ist es in den meisten Fällen so, daß „Erfaßt-sein-durch-Metarepräsentation" und „Eingebettetsein-in-ein-Selbstmodell" als Eigenschaften einer bestimmten Klasse mentaler Modelle *koinstantiiert* sind. Diese Tatsache hat auf philosophischer Ebene häufig zu vorschnellen begrifflichen Identifikationen geführt. Die Schlafwandlerin, die im hellen Mondschein mit großer Sicherheit über den Dachfirst spaziert, ist vielleicht ein Beispiel für ein System, das ein sehr komplexes und funktional aktives Selbstmodell erzeugt, ohne die in dieses Selbstmodell eingebetteten Zustände noch einmal metarepräsentational abzubilden. Phänomenale Meinigkeit ist allerdings an Bewußtheit gebunden: Es gibt in solchen Fällen kein subjektives *Erleben* der eigenen Identität.

> **P 11**: *Ist die Subjektzentriertheit mentaler Zustände eine mit ihrem Erlebnischarakter notwendig verknüpfte Eigenschaft oder kann es Subjektivität ohne qualitativen Gehalt geben?*

Hier gilt *mutatis mutandis*, was ich über die Mathematiker im Mathematikerhimmel gesagt habe: Selbstmodellerzeuger ohne Analog-Indikatoren sind durchaus denkbar. In unserem eigenen Fall werden immer auch mentale Präsentate aktiviert, während Selbstmodellierung stattfindet. Das ist so, weil das repräsentationale Fundament unseres Selbstmodells, das Körperschema, selbst mit propriozeptiven Präsentaten unterlegt ist, denen auf der phänomenalen Ebene das Körpergefühl und die Schwereempfindungen entsprechen. Die neurobiologische Grundlage des repräsentationalen Fundaments wiederum könnte die kontinuierliche, inputunabhängige Aktivität einer genetisch fixierten neuronalen *Körpermatrix* sein.[33] Wir sind deshalb Wesen, welche sich in repräsentationalen Standardsituationen immer

[33] Vgl. Melzack 1989, 1992, Melzack *et al.* 1997.

schon als konkrete und verkörperte Systeme erleben, also als Wesen, die sich zumindest teilweise einer scheinbar direkten *Selbstgegebenheit* erfreuen.

Wenden wir uns nun noch einmal kurz der erkenntnistheoretischen Problemdimension zu und erinnern uns an die folgenden, eingangs gestellten Fragen:

> **E 1**: *Ist „Subjektivität" eine epistemische Relation? Können mentale Repräsentate – ähnlich wie Sätze – als Wahrheitswertträger fungieren?*

In meinen einführenden Bemerkungen zu der hier zugrundegelegten Theorie über die Entstehungsbedingungen von mentalem Gehalt – sie wurde im zweiten Kapitel in Anlehnung an McGinn auch als „Cambridge-Theorie mentaler Repräsentation" bezeichnet – habe ich bereits darauf hingewiesen, daß die empirisch plausiblen Entitäten *mentale Modelle* sind – also Analogrepräsentate, die keine Wahrheitswerte besitzen und untereinander nicht in Folgerelationen stehen. Mentale Modellierung benutzt abstrakte Isomorphismen, um Ähnlichkeitsbeziehungen zu ihren Repräsentanda herzustellen. Die Erforschung dieser höherstufigen Isomorphismen ist möglicherweise das Kernproblem für eine Naturalisierung der Erkenntnistheorie. Innere Zustände ähneln ihren Objekten, aber sie referieren in unserem Fall nicht auf sie. Natürlich sind auch Systeme denkbar, die dem Fodorschen Traum einer *Language of Thought* gerecht werden, die hier vorausgesetzten mentalen Modelle und auch Selbstmodelle erzeugen aber kein digitales Wissen, sondern ein analoges Wissen. Subjektivität im Sinne der Eingebettetheit eines mentalen Modells in ein Selbstmodell ist demnach eine Beziehung zwischen zwei internen Repräsentaten. Über diese repräsentationale Relation kann das System eine bestimmte Klasse von Fakten mental simulieren und repräsentieren: Nämlich solche Fakten, die darauf beruhen, daß eine bestimmte Teilmenge von Weltzuständen gleichzeitig *Systemzustände* sind.

> **E 6**: *Ist in der „Meinigkeit", d. h. der erlebnismäßigen Subjektzentriertheit mentaler Zustände eine Form von Wissen kodiert?*

„Meinigkeit" ist eine phänomenale Eigenschaft, die durch die eben angesprochene repräsentationale Einbettungsrelation ins Spiel kommt. Über die Herstellung dieser Relation in seinem mentalen Modell der Realität erwirbt ein System analoges Wissen, indem es nun eine bestimmte Klasse von Fakten intern abbilden kann. Der epistemische Agent ist dabei immer das System, nicht das psychologische Subjekt. Allerdings ist die innere Erfahrung der „Meinigkeit" eine *Widerspiegelung* des epistemischen Prozesses: Die erlebnismäßige Subjektzentriertheit mentaler Zustände ist nicht identisch mit dem Vorgang der neuronalen Informationsverarbeitung, durch den die entsprechenden mentalen Modelle aktiviert werden. Son-

dern sie ist vielmehr die Art und Weise, in der sich das System diesen Vorgang, seine Vorbedingungen und seine Resultate auf der Ebene phänomenaler Zustände selbst zu erklären versucht.

> **E 4**: *Was bedeutet es, daß subjektive mentale Zustände „private" Zustände sind?*

Subjektive mentale Zustände sind Zustände von mentalen Modellen, die in das Selbstmodell eines Systems eingebettet wurden. Trivialerweise sind solche Zustände in einem simplen physikalischen Sinn Zustände, die einer Person oder einem System gehören, weil sie *ihre* Zustände sind.[34] Es sind also nicht private Objekte oder phänomenale Individuen, zu denen das psychologische Subjekt in einer exklusiven Besitzbeziehung steht, sondern variable interne Datenstrukturen, über deren angebliche Privatheit wir zu entscheiden haben. Das psychologische Subjekt ist ja seinerseits eine solche durch mentale Repräsentation erzeugte Datenstruktur: Das mentale Modell des Systems *selbst*. Wenn mit der problematischen „Privatheit" mentaler Zustände, die in verschiedenen Varianten von Philosophen immer wieder als Kandidat für ein kategoriales Abgrenzungskriterium mentaler Zustände gegenüber physischen Zuständen diskutiert wurde, nicht die physikalische Internalität von Systemzuständen gemeint sein soll, kann sie somit noch als *repräsentationale* Internalität interpretiert werden. Die Tatsache, daß zwischen gewissen internen Datenstrukturen eine Einbettungsrelation besteht, kann aber ihrerseits schwer zu einem metaphysischen Mysterium oder zu einem interessanten Einwand gegen eine naturalistische Theorie des Subjekts umgemünzt werden. In diesem Kontext ist es nicht mehr als ein schlichtes, kontingentes Faktum, daß manche auf natürliche Weise entstandenen Repräsentationssysteme einige der von ihnen intern aktivierten mentalen Modelle noch einmal in ein Selbstmodell einbetten.

Also könnte ein auf dem Begriff der „Privatheit" beruhender Einwand gegen die Naturalisierung des psychologischen Subjekts letztlich nur ein *phänomenologischer* sein, der die phänomenale „Meinigkeit" subjektiver Zustände zu seinem Ausgangspunkt macht. Das kann man tun, indem man die *phänomenale* Internalität – also die psychische Erfahrung von Innerlichkeit und Perspektivität – zu einem *Evidenzerlebnis* überhöht. Man könnte dann versuchen, aus dieser inneren Evidenz ein Argument für die *epistemische Geschlossenheit des inneren Raumes* zu konstruieren. Dazu braucht man aber eine starke erkenntnistheoretische Prämisse, die ich häufig als die cartesianische Annahme der *Selbsttransparenz des Subjekts* bezeichnet habe. *Diese* Annahme ist jedoch angesichts des neueren empiri-

[34] Vgl. hierzu und bezüglich des Scheiterns der Inkorrigibilitätsthese als Einwand gegen die Identitätstheorie Bieri 1981: 60; Metzinger 1985, Abschnitte 2.1.7.1 und 2.1.7.2, außerdem Rorty 1981a, 1981b (1965, 1970): 93, besonders 107ff. Eine Zusammenstellung der kanonischen Texte der Privatheits-Diskussion findet sich in Bieri 1981: 344f. Vgl. ferner Wilkes 1984: 225.

schen Materials über die neurobiologischen und neuroinformatischen Hintergrundbedingungen phänomenalen Selbstbewußtseins beim Menschen unhaltbar geworden.

Es gibt allerdings einen starken Sinn, in dem das menschliche Gehirn einen privilegierten Zugang zu den von ihm aktivierten mentalen Modellen hat. Es erfaßt bereits jetzt durch Metarepräsentation viele der interessanten Eigenschaften der von ihm erzeugten Datenstrukturen, Eigenschaften, die die Neurowissenschaftler aus ihrer externen Perspektive vielleicht niemals oder erst in ferner Zukunft werden entdecken und beschreiben können. Wenn zum Beispiel die hier vorgeschlagene Hypothese über mentale Präsentate in die richtige Richtung deutet, dann gibt es abstrakte Eigenschaften solcher Datenstrukturen – nämlich *Mikroformate* –, die vom Gehirn zielsicher erfaßt und in Gestalt derjenigen Zustände dargestellt werden, die wir dann als Qualia bezeichnen. Das Gehirn ist gegenüber dem Wissenschaftler, der diese abstrakten Eigenschaften verstehen will, als das diese Eigenschaften erzeugende System in einer privilegierten Position. Mentale Modelle sind, wenn wir sie als Aktivierungsvektoren in den inneren Schichten eines komplexen und massiv parallel arbeitenden Systems analysieren, abstrakte Entitäten, die Transformationen in hochdimensionalen Vektorräumen durchlaufen. Gegenüber dem Wissenschaftler hat das Gehirn den Vorteil, diese abstrakten Prozesse metarepräsentational *als Konkreta* darstellen zu können. Vielleicht ist unser phänomenales Bewußtsein am Ende doch keine so schlechte Darstellung der inneren Prozesse, vielleicht bündelt diese *subjektive* Repräsentation in analogem Format letztlich doch einen Großteil der auch aus theoretischer Perspektive interessanten Eigenschaften des Systems. Unser phänomenales Bewußtsein inkorporiert schließlich eine große Menge nicht-diskursiven Wissens über uns selbst und die Vergangenheit unserer Spezies auf diesem Planeten – nur darf man nicht übersehen, daß die Erzeugung dieses Wissens sehr spezifischen Interessen diente, die durch biologische Imperative determiniert wurden. Wo diese Imperative nicht gelten, ist es hochgradig fallibel.

LS 1: *Ist das Subjekt logisch primitiv?*

LS 5: *Funktionieren mentalistische Terme wie Eigennamen, d. h. spezifizieren sie ihre Referenten über alle logischen Welten hinweg?*

Ich habe dafür argumentiert, daß das, was wir gemeinhin als das psychologische Subjekt zu bezeichnen pflegen, substituierbar ist durch eine Entität innerhalb einer empirisch verankerbaren Theorie mentaler Repräsentation, die ich das *mentale Selbstmodell* eines Systems genannt habe: Psychologie ist im Kern die wissenschaftliche Erforschung einer bestimmten Datenstruktur, wenn man so will *phänomenale Informatik*. Über die logische Grundstruktur unserer Ontologie im allgemeinen ist damit natürlich nichts ausgesagt.

Allerdings impliziert die Selbstmodell-Theorie der Subjektivität eine Vereinfachung der psychologischen Ontologie. Sie sagt, daß Sätze über Zustände des psychologischen Subjekts oder des phänomenalen Selbst letztlich Sätze über eine variable Datenstruktur innerhalb des beschriebenen Systems sind. Psychologische Aussagen handeln – ganz gleich, ob sie in der ersten oder in der dritten Person formuliert werden – von einer bestimmten Klasse mentaler Modelle. Das heißt: Für alle psychologischen Belange können wir die Annahme eines unveränderlichen Wesenskerns aufgeben, das *Ego* als ein Individuum innerhalb der mentalistischen Ontologie kann eliminiert werden. Die Zentrierung interner Simulationsräume und die Perspektivität phänomenaler Zustände können durch mentale Selbstmodellierung befriedigend erklärt werden – zumindest gibt es gute Gründe zu der Annahme, daß solche Erklärungen zukünftig möglich sein werden. Das Ich ist keine primitive, irreduzible Entität der psychologischen Beschreibungsebenen[35] mehr, weil sich Möglichkeiten zu eröffnen beginnen, diese Entität auf begrifflich präzise und empirisch gehaltvolle Entitäten darunterliegender *repräsentationaler* Erklärungsebenen zu reduzieren.

Es ist mir jedoch in diesem Buch nicht darum gegangen, eine reduktionistische These zu entwickeln oder eine strikte *Identifikation* repräsentationaler und phänomenaler Entitäten nahezulegen. Für ein solches Unternehmen ist es noch zu früh, es kann beim gegenwärtigen Stand unseres empirischen Wissens nur zur Spekulation verkommen. Klassische Identitätsthesen – etwa im Sinne von Place, Smart oder Armstrong durch Sätze des Typs *„Das Selbstmodell ist das phänomenale Selbst“* oder *„Was ein mentales Modell zu einem subjektiven Zustand macht ist die Einbettung in ein Selbstmodell“* – wären als kontingente Aussagen derzeit schlicht uninteressant, weil ihr empirischer Gehalt nur sehr schwach ist. Solche Thesen wären nicht *informativ*. Mein Ziel ist deshalb lediglich gewesen, Wege aufzuzeigen – Wege in Richtung auf eine erfolgversprechende begriffliche Integration der unter dem Begriff „Subjektivität mentaler Zustände“ gebündelten psychologischen Eigenschaften in eine naturalistische Theorie des Geistes, die für zukünftige Entwicklungen offen ist.

Was die essentialistische Interpretation mentalistischer Terme als *rigide Designatoren* angeht (vgl. Abschnitt 1.1), kann man sagen: Das phänomenale Erlebnis der Meinigkeit, die das Kripkesche Argument gegen die Identitätstheorie ausbeutet, indem sie die durch diese Eigenschaft mancher innerer Erlebnisse begründeten Intuitionen semantisch überhöht, hat ihre Instantiierungsbasis in einer kontingenten Einbettungsrelation zwischen mentalen Modellen. Wenn nun jemand einwendet: *„Wer behauptet, es könnte Schmerzen oder andere mentale Zustände geben, die* ***niemandes*** *Schmerzen oder mentale Zustände sind, der hat nicht verstanden, was*

[35] Eine bedeutsame philosophische Frage, der ich hier nicht nachgehen kann, ist die nach der Notwendigkeit eines primitiven psychologischen Subjekts für die Aufrechterhaltung der *personalen* Beschreibungsebene. Gehört die Annahme eines stabilen psychischen „Wesenskerns“ notwendig zum Begriff der „Person“?

Schmerzen sind!", kann man darauf erwidern: *„Der hier postulierte Begriff des **Mentalen** ist empirisch unplausibel. Wer einen essentialistischen Begriff der Subjektgebundenheit mentaler Zustände voraussetzt, kann viele Fälle devianter mentaler Selbstmodellierung und andere neuere Befunde nicht mehr adäquat interpretieren und hat deshalb **seinerseits** nur ein sehr begrenztes Verständnis davon, was Schmerzen und andere mentale Zustände in Wirklichkeit sind."* Es gibt mögliche Welten, in denen das Gehirn die Einbettung mentaler Modelle in das aktuelle Selbstmodell nicht mehr leisten kann, zum Beispiel wenn die betreffende Person einen schizophrenen Schub erleidet. Dann sind aktive mentale Modelle von höheren kognitiven Operationen möglicherweise auf der phänomenalen Ebene nicht mehr *meine Gedanken*, sondern *fremde Gedanken*, die meinen inneren Erlebnisraum penetrieren. Eine moderne Theorie des Mentalen muß auch solche und andere Fälle *zerfallender* Subjektivität erklären können – mit einem essentialistisch-aprioristischen Begriff des psychologischen Subjekts kann ihr dies nicht gelingen.

> **P 12**: *Welche Klasse von Systemen kann die unter dem Begriff „Subjektivität" gebündelten Eigenschaften instantiieren? Ist künstliche Subjektivität möglich?*

Ich habe die interessante Klasse von Systemen über ihre repräsentationalen Merkmale provisorisch auszuzeichnen versucht, weil ich der Meinung bin, daß das psychologische Phänomen „Subjektivität" im Rahmen einer repräsentationalen Theorie des Geistes erschöpfend analysiert werden kann. Die Subjektivität im starken Sinne eines qualitativen, phänomenalen Selbstbewußtseins instantiierenden Systeme sind demnach Selbstmodellerzeuger, die einen Teil ihrer mentalen Präsentate und ihres Selbstmodells durch eine einheitliche metarepräsentationale Funktion noch einmal intern abbilden. Ich möchte diesen Punkt an zwei häufig vorgebrachten Einwänden illustrieren.

Einwand: *Mit „Selbstmodellierung" greifen Sie ein für die Erklärung von Subjektivität irrelevantes Kriterium heraus. Wir könnten jederzeit eine Maschine bauen, die ein Selbstmodell erzeugt und in ihre innere Darstellung der Welt integriert. Niemand würde glauben, daß ein solches System Bewußtsein besitzt!*

Antwort: Richtig. *Bewußte Subjektivität* entsteht dadurch, daß ein System zumindest einen Teil seines Selbstmodells noch einmal durch eine vereinheitlichende Metarepräsentationsfunktion der höchsten Ebene in der repräsentationalen Hierarchie (wie ich sie in Abschnitt 2.1.4 hypothetisch vorgeschlagen habe) intern abbildet. Wir müssen also **SMT** und **MRT** kombinieren:

(SMT, MRT): Subjektivität ist eine Eigenschaft komplexer informationsverarbeitender Systeme, die genau dann instantiiert wird, wenn das System in das von ihm aktivierte Realitätsmodell ein *Selbstmodell* einbettet. Die Inhalte phänomenalen Bewußtseins sind Meta-Repräsentate, die für ein System eine Teilmenge der gegenwärtig in ihm aktivierten mentalen Simulate und Präsentate abbilden. Das, was alle bewußten inneren Zuständen miteinander gemein haben, ist die Tatsache, daß sie durch eine globale *Metarepräsentationsfunktion erfaßt* werden. *Bewußte Subjektivität* entsteht, wenn ein Teil des Selbstmodells durch diese Funktion erfaßt wird.

Die Kombination von **SMT** und **MRT** bietet den begrifflichen Rahmen für eine Analyse bewußter Subjektivität im Rahmen einer naturalistischen Theorie mentaler Modellierung: Bewußte Subjektivität instantiiert ein System genau dann, wenn es einen Teil des von ihm aktivierten Selbstmodells metarepräsentational abbildet.

Einwand: *Gottseidank sind Maschinen, die dieses Merkmal besitzen und im Komplexitätsbereich des Menschen operieren, technisch noch nicht machbar. Aber selbst, wenn es sie einmal geben sollte, hätten sie keine* ***qualitativ*** *gehaltvollen inneren Zustände – die Rotwahrnehmungen einer solchen Maschine würden ohne die Instantiierung* ***phänomenaler Röte,*** *die internen Repräsentationen von Beschädigungen ihrer Hardware ohne* ***Schmerzhaftigkeit*** *in „erlebnismäßiger Blindheit" ablaufen. Das Nagelsche „****Wie es ist, ein X zu sein****" gäbe es für diese Maschine nicht.*

Antwort: Richtig. Wenn ein solches System aber das erzeugt, was ich als *mentale Präsentate* bezeichnet habe, und dann noch in den metarepräsentational abgebildeten Teil des Realitätsmodells einbettet, erfüllt es die repräsentationalen Bedingungen für die Erzeugung von denjenigen Zuständen, die wir in der – etwas vagen und möglicherweise begrifflich inkonsistenten – traditionellen Taxonomie als „Qualia" bezeichnen. Wir müssen also für die Analyse qualitativen Selbstbewußtseins noch eine Zusatzhypothese aufstellen. Ich kann hier nur die *Analog-Indikator-Hypothese* zur Adjunktion anbieten:

(SMT, MRT, AIT): Subjektivität ist eine Eigenschaft komplexer informationsverarbeitender Systeme, die genau dann instantiiert wird, wenn das System in das von ihm aktivierte Realitätsmodell ein *Selbstmodell* einbettet. Die Inhalte phänomenalen Bewußtseins sind Meta-Repräsentate, die für ein System eine Teilmenge der gegenwärtig in ihm aktivierten mentalen Simulate und Präsentate abbilden. Das was alle bewußten inneren Zuständen miteinander gemein haben, ist die Tatsache, daß sie durch eine globale *Metarepräsentationsfunktion* erfaßt werden. Qualia sind *Analog-Indikatoren*, die für ein System die aktuelle Gegenwart eines Präsentandums intern signalisieren. Das, was alle introspektiv erlebbaren mentalen Präsentate (die durch den Vorgang erzeugten aktiven Datenstrukturen) eines phänomenalen Typs (eines Quale) miteinander gemein haben, ist ihr *Format*.

Dies ist die These der Selbstmodell-Theorie in ihrer umfassenden Formulierung: *Qualitatives, subjektives Bewußtsein* entsteht durch die Aktivierung eines mit mentalen Präsentaten unterlegten Selbstmodells, das von einer Metarepräsentationsfunktion der höchsten Stufe erfaßt wird. Man muß dabei immer betonen, daß dies keine dogmatische Subjekttheorie sein soll, sondern der plastische begriffliche Rahmen für zukünftige – empirisch immer weiter angereicherte und auch durch Präzisierungen der zugrundegelegten Theorie mentaler Repräsentation verbesserte – naturalistische Konzeptionen des psychologischen Subjekts. Es gibt also nach der hier skizzierten Konzeption verschieden starke Formen von Subjektivität. Selbstmodellierung alleine erzeugt nur Subjektivität in einem *schwachen* Sinne.

Einwand: *Das ist ja alles gut und schön, meinetwegen können Sie dann statt Künstliche-Intelligenz-Forschung auch noch Künstliche-Subjektivitäts-Forschung betreiben. Ich glaube jedoch nicht, daß jemals die technologische Realisierung eines künstlichen Subjekts gelingen wird. Ich muß aber auf der anderen Seite zugeben, daß meine Intuitionen bezüglich der von Ihnen zuletzt definierten Systemklasse nicht mehr so eindeutig sind – letztlich steht hier Intuition gegen Intuition. Trotzdem wird ein künstliches System, das der durch die Kombination von* **SMT, MRT** *und* **AIT** *gekennzeichneten Klasse angehört, niemals* ***Gefühle*** *haben: Vielleicht erzeugt es sensorische Qualia,* ***emotionale Qualia*** *besäße es jedoch nicht. Und die sind fundamental für* ***echte*** *Subjektivität.*

Antwort: Was Sie als die „authentische" Subjektivität immer wieder suchen, könnte am Ende ein essentialistisches anthropozentrisches Vorurteil sein. Sie haben jedoch recht: Um *emotionale* mentale Präsentate zu aktivieren, müßte ein künstliches System genuine Interessen besitzen, die sich nicht von den Interessen seiner Konstrukteure ableiten. Dieses Kriterium kann ein System aber letztlich nur erfüllen, wenn es auch ein *soziales*

Subjekt ist. Es müßte sich in einer physischen Wettbewerbssituation mit anderen Systemen befinden, etwa um Energiequellen, Verbesserung seines sozialen Status oder um die erfolgreiche Replikation seiner funktionalen Struktur auf anderen Trägersystemen. Außerdem müßte es in der Lage sein, Operationen auf einer intern repräsentierten Bedürfnishierarchie durchzuführen. Das heißt: Es müßte genuine, selbst-definierte Meta-Ziele erzeugen und mental modellieren, es müßte das besitzen, was Harry Frankfurt als *„second-order-volitions"* bezeichnet hat.[36] Dann könnten allerdings innere Zustände entstehen, die die aktuelle Interessenlage des Systems auf ähnliche Weise phänomenal darstellen wie wir selbst. Ein künstliches Subjekt könnte man also nur im Rahmen einer *künstlichen Gesellschaft* erzeugen. Die *kompetitive* Integration künstlicher Systeme in menschliche Gesellschaften wäre allerdings ein Schritt, den man sorgfältig überlegen sollte: Wir könnten – und das ist ja auch das zentrale intuitive Element des populären Ressentiments gegen die KI-Forschung – zu hilflosen Zauberlehrlingen werden, die eine postbiotische Evolution künstlicher Subjekte in Gang gesetzt haben, die sie nun schon allein aus ethischen Gründen nicht mehr unterbrechen dürfen.

Die Realität wird solche Phantasien so bald nicht einholen. Die Erfahrung hat bereits gezeigt, daß der Teufel in technischen Details steckt und nicht in theoretischen Modellen. Es könnte sein, daß es einfach keine nicht-biologische *Hardware* gibt, die das funktionale Profil der für die Instantiierung eines subjektzentrierten, phänomenalen Bewußtseins erforderlichen repräsentationalen Abläufe physikalisch realisieren kann. Außerdem gibt es eine Reihe weiterer Kriterien, die jedes System erfüllen müßte, wenn es ein Subjekt im starken Sinne sein soll:

– **Phänomenale Historizität:** Ein solches System muß eine *Geschichte* besitzen (eine „Kindheit", in der es seine eigene innere Konfiguration in Interaktion mit der Umwelt herausbildet). Diese Geschichte muß durch mentale Selbstmodellierung intern abgebildet werden, so daß eine *subjektive Biographie* entsteht und das System seine eigene Geschichtlichkeit, seine innere und äußere Prozessualität auch als seine *eigene* bewußt erleben kann.

– **Implizites Hintergrundwissen**: Das ganz spezifische „In-der-Welt-Sein" menschlicher Subjekte hängt auch mit dem enormen Umfang und der Komplexität ihres *impliziten* Wissens über die Struktur ihrer sozialen Umgebung und der evolutionsbiologisch relevanten Aspekte der physikalischen Umwelt zusammen. Diese Tatsache ist häufig und meiner Auffassung nach berechtigterweise als Argument gegen die starke KI-These bezüglich klassischer KI-Systeme mit einer sequentiellen Architektur ins Feld geführt worden[37]. Ob sie sich auch erfolgreich gegen neuronale Netze oder Hybrid-Systeme wird einsetzen lassen, bleibt abzuwarten.

---

[36] Vgl. Frankfurt 1981 (1971).

[37] Vgl. etwa Dreyfus 1985, Dreyfus/Dreyfus 1987 und die in Born 1987 zusammengestellten Texte.

– **„Propositionale Subjektivität“**: Von einem überzeugenden künstlichen Subjekt müßten wir außerdem mindestens verlangen, daß es eine postcartesianische *res cogitans* ist: Ein physikalisches System, das propositionalen Gehalt mental modellieren kann, und zwar indem es die mentalen Modelle externer, satzartiger Strukturen und der mit ihnen möglichen Operationen gleichzeitig in sein Selbstmodell einbettet. Nur so könnte ein *phänomenales Cogito* entstehen und damit ein psychologisches Subjekt, das sich als ein denkendes Subjekt erleben kann. Wir selbst besitzen diese Fähigkeit nur in sehr beschränktem Ausmaß: Die mentale Modellierung einer noch relativ simplen Operation in einem digitalen Code – zum Beispiel das Lösen der Aufgabe 4179 x 2968 „im Kopf“ – überfordert unsere Gehirne sehr schnell. Mit externen Instrumenten wie Papier, Bleistift und den mit ihnen erzeugbaren Schriftzeichen können wir diese Aufgabe in einer Sequenz von *äußeren* Schritten jedoch leicht lösen (unter anderem deshalb, weil wir nicht mit der Tatsache zu kämpfen haben, daß die mentalen Repräsentate der Zwischenschritte nach wenigen Sekunden aus dem Puffer unseres Kurzzeitgedächtnisses verschwinden). Kurz: Nur ein in eine Sprechergemeinschaft integriertes künstliches System könnte eine externe Sprache semantisch zu beherrschen lernen. Nur ein solches System wäre auch in der Lage, die durch eine Sprache ermöglichten kognitiven Leistungen dann intern noch einmal zu modellieren.

Ich habe eine repräsentationale Analyse des psychologischen Subjekts angeboten. Aus der hier vorgeschlagenen Selbstmodell-Theorie geht nicht hervor, welche Klasse von physikalischen Systemen die betreffenden *„bottom-up constraints“* erfüllen und deshalb qualitativ gehaltvolles Selbstbewußtsein instantiieren können. Im Rahmen der „starken“ Selbstmodell-Theorie (im Sinne einer Kombination der Hypothesen **SMT**, **MRT** und **AIT**) ist es also durchaus möglich, daß es einmal nicht-biologische Subjekte geben wird.[38] Die ersten Forschungsprogramme mit dem Entwicklungsziel „introspektiver KI-Systeme“[39] haben bereits begonnen. Trotzdem sollten wir uns diesbezüglich vor einem naiven und vorschnellen Optimismus (oder Pessimismus) hüten. Wie wir eben gesehen haben, gibt es interessante Kandidaten für *zusätzliche* Kriterien, die ein artifizielles System erfüllen muß, wenn wir es als genuines psychologisches Subjekt in einem weiteren Sinne akzeptieren sollen. Einen letzten normativen Charakter werden solche Kriterien allerdings nicht verlieren. Und es wäre unvorsichtig, von der

[38] Mentale Modelle sind, wenn sie nur über ihren Gehalt und nicht über ihr funktionales Profil oder ihre spezifischen Aktivierungsbedingungen beschrieben werden, *ontologisch neutrale Entitäten*. Hilary Putnam hat bereits Mitte der sechziger Jahre darauf hingewiesen, daß auch Engel Turing-Maschinen sein könnten, weil der klassische Maschinen-Funktionalismus nichts über die Art aussagt, in der ein abstrakter Automat letztlich realisiert wird. Auch Selbstmodellgeneratoren könnten theoretisch auf nicht-physikalischer „Astralmaterie“ oder auf *„angel stuff“* realisiert sein und in diesem Sinne nicht-biologische Subjekte instantiieren. Die teleofunktionalistische Analyse solcher Systeme bekäme aber Schwierigkeiten bei der Spezifizierung von Input-Output-Relationen und Zielvariablen. Vgl. Putnam 1975.

[39] Einen Überblick über Bereiche der Metakognitions- und Selbstkonzeptforschung gibt Kiefer 1988, Teil I.

Annahme auszugehen, daß zukünftige Versionen einer naturalistischen Theorie des Geistes nicht ganz neue Kriterien enthalten werden. Das beste und befriedigendste Kriterium für die Anerkennung eines nicht-biologischen Subjekts, das ich persönlich kenne und favorisieren würde, wäre genau dann erfüllt, wenn ein System uns überzeugend demonstriert, daß es das in dieser Untersuchung diskutierte *philosophische Problem* der Subjektivität mentaler Zustände verstanden hat und mit eigenen Argumenten in die Diskussion einzugreifen beginnt.

Ich habe im vierten Kapitel die skeptischen Nagelschen Argumente gegen die Möglichkeit einer erschöpfenden objektiven Theorie des Mentalen zu einem tieferen kommentierenden Eindringen in die Problematik benutzt. Nagels Philosophie des Subjekts erschien mir unter anderem deswegen wertvoll, weil sie sich auf einem hohen analytischen Niveau bewegt und erfolgreich der Ignorierung oder Einebnung einer zentralen Problemdimension durch die in der neueren angelsächsischen Diskussion teilweise anzutreffende reduktionistische Euphorie[40] entgegengewirkt hat. Wenn die phänomenale Perspektivität unseres Innenlebens mittlerweile ein etablierter Bestandteil der Diskussion ist – der sie wiederum attraktiver für kontinentale Denker macht –, dann ist das auch ein Verdienst von Thomas Nagel. Ich werde deshalb diesen Abschnitt vorläufiger Antworten beenden, indem ich zu Nagels Formulierungen des Problems zurückkehre.

> **E 7**: *Läßt sich aus der Subjektivität mentaler Zustände die These ableiten, daß das wissenschaftliche Weltbild prinzipiell unvollständig bleiben muß?*

Die wissenschaftliche Repräsentation der Welt durch sich entwickelnde Theorien ist ihrer Struktur nach *aperspektivisch*: In dieser Repräsentation wird das eben jene Repräsentation erzeugende System, die *scientific community*, nur unvollständig „systemintern" dargestellt (zum Beispiel in den wiederum von ihr selbst erzeugten wissenssoziologischen und wissenschaftstheoretischen Beschreibungen derselben). Die Inhalte des durch den Prozeß der wissenschaftlichen Theoriebildung geöffneten Repräsentationsraumes sind intersubjektiv überprüfbar und gegebenenfalls auch intersubjektiv gültig. Darin liegt ihre Stärke: Sie sind keine *benutzerfixierten* Repräsentate.

Im Fall der Humanpsychologie des Wachbewußtseins handeln sie aber notgedrungen von mentalen Modellen der Wirklichkeit, die keine Variablen enthalten und durch die Einbettung eines Selbstmodells zu phänomenal zentrierten und von der funktionalen Analyse her *benutzerfixierten* Repräsentaten geworden sind. Selbst wenn wir durch eine teuflische Technologie die zentrierten Realitätsmodelle zwischen zwei Systemen austau-

[40] Karl Popper hat diese metaphysische Ideologie treffend als *„promissive materialism"* bezeichnet, als eine Versprechungen machende Haltung, die Schuldscheine auf zukünftige theoretische und empirische Entwicklungen aufnimmt. Vgl. Popper/Eccles 1982.

schen könnten, würden sie aus der Perspektive einer teleofunktionalistischen Analyse zwar auch weiterhin als sehr komplexe interne Systemzustände erscheinen, aber nicht mehr als solche, die eine Funktion *für* ihren Benutzer, *für* das sie aktivierende System haben. Diese Benutzerfixierung durch ein von Variablen freies Selbstmodell dient als Instantiierungsbasis eines spezifischen Sets von psychologischen Eigenschaften. Es gibt keinen prinzipiellen Grund zu der Annahme, daß erfolgreiche wissenschaftliche Beschreibungen der repräsentationalen, funktionalen und neurobiologischen Merkmale dieser Instantiierungsbasis psychologischer Subjekte zukünftig nicht Teil des wissenschaftlichen Weltbildes werden könnten.

Es ist nun trivial wahr, daß solche Beschreibungen *per se* nicht die fraglichen psychologischen Eigenschaften in die Welt bringen und es ist außerdem sehr wahrscheinlich, daß solche Beschreibungen *allein* unsere Fähigkeiten zur mentalen Simulation von Fremdsystemen (z. B. Fledermäusen) nicht wesentlich verbessern werden. Auch wenn uns die technologische Umsetzung einer an ihr historisches Ende gekommenen Theorie des Geistes gelänge und wir künstliche Subjekte erzeugen könnten, dann entstünden diese künstlichen Subjekte der hier vertretenen Theorie zufolge ja immer durch die Aktivierung variablenfreier benutzerzentrierter Selbstmodelle. Sogar wenn es uns gelänge, künstliche oder kopierte Selbstmodelle innerhalb der mentalen Simulationsräume lebender Menschen zu implantieren, könnte es zwar zu interessanten inneren Dialogen kommen (die den erkenntnistheoretischen Disputen mit Traumfiguren in einem Klartraum ähneln würden), aber niemals zu der Nagelschen Qualität des *„Wie es ist, ein X zu **sein**“*. Entweder wir würden in den bedauernswerten Versuchspersonen eine Episode multizentrierten Bewußtseins nach Art der bereits geschilderten *Multiple Personality Disorders* erzeugen. Oder das Experiment zöge einen Zusammenbruch des ursprünglichen psychologischen Subjekts mit einer darauf folgenden Amnesie für den betreffenden Zeitraum nach sich.

Durch ein von Variablen freies Selbstmodell zentrierte mentale Modelle der Realität können also aus einem prinzipiellen Grund nicht vollständig dupliziert werden: Wegen ihrer funktional-repräsentationalen *Fixiertheit* auf das sie erzeugende und benutzende System. Aus diesem Grund werden wir auch niemals wissen, wie es ist, eine Fledermaus zu **sein**. Und aus eben demselben Grund sind wir auch *als geistige Wesen* noch in einem sehr starken Sinn körpergebunden: Wir können die biologischen Systeme, die uns erzeugen, auch als Subjekte phänomenaler Zustände nicht transzendieren.

> **P 13**: *Gibt es für externe Beobachter eine zuverlässige Möglichkeit, um zu entscheiden, ob in einem gegebenen System mit bekannten physischen und funktionalen Eigenschaften die phänomenale Qualität des Nagelschen „Wie es ist, ein X zu sein“ auftritt?*

Wenn die repräsentationale Analyse von Qualia als mentalen Präsentaten und ihre funktionale Analyse als *für* ein System arbeitenden Analog-Indika-

toren eine erfolgversprechende Strategie darstellt, dann gibt es zumindest prinzipiell die Möglichkeit, die fraglichen abstrakten Eigenschaften der durch die betreffenden funktionalen Prozesse erzeugten Datenstrukturen (die ich als ihr „Format" bezeichnet habe) auch aus systemexterner Perspektive dingfest zu machen. Das ist natürlich nicht gleichbedeutend mit einer Erzeugung der entsprechenden Qualia im Bewußtsein der jeweiligen Forscher: Wir können bestenfalls digitales Wissen über mentale Präsentate und die durch sie instantiierten phänomenalen Zustände erlangen, jedoch kein subjektives analoges Wissen. In *diesem* Sinne werden wir niemals wissen, wie es ist, eine Fledermaus zu sein.

Der zentrale Aspekt psychologischer Subjektivität kommt aber, wie wir gesehen haben, durch Selbstmodellierung ins Spiel. Wie können wir – angenommen, daß wir die internen Eigenschaften eines Systems nicht oder nur unvollständig kennen – herausfinden, ob es ein mentales Selbstmodell besitzt? Wir müssen untersuchen, ob es auf der Ebene äußeren Verhaltens Leistungen zeigt, deren Vorbedingung ein funktional aktives Selbstmodell ist.

Auf der höchsten Ebene, der Ebene *linguistischer* Kapazitäten, kann man fragen: Verfügt das System über ein Selbst-Symbol, mit dem es in sprachlichen Kommunikationssituationen auf sich selbst Bezug nehmen kann? Soziale Selbstbezugnahme ist eine biologisch sinnvolle Fähigkeit. Wenn Systeme beginnen, in externen Codes auf sich selbst Bezug zu nehmen, dann ist dies ein starkes Indiz dafür, daß sie bereits über *interne* Formen der Selbstreferenz bzw. Selbstmodellierung verfügen. Auf einer niedrigeren Stufe, der Stufe reflexiver Deixis, kann man untersuchen, ob ein gegebenes System – etwa in einer Gruppe – *auf sich selbst zeigen kann*. Wie könnte es dies tun, ohne zumindest ein rudimentäres, funktional aktives Selbstmodell aktiviert zu haben? Und schließlich kann man das Vorhandensein der Fähigkeit zur Selbstmodellierung nach den einzelnen jeweils vorhandenen *Sinnesmodulen* überprüfen: Kann ein System seinen eigenen Körper zuverlässig ertasten, hat es ein Gespür für seine *Körpergrenzen*? Erkennt es seine eigenen *Geräusche* wieder, zum Beispiel auf Tonbandaufnahmen? Kann es sich selbst *olfaktorisch* repräsentieren, indem es – wie ein Hund, der seine Duftmarken wiedererkennt – von ihm verursachte Geruchseigenschaften der Welt *als seine eigenen entdeckt?* Ist es sich auch über den visuellen Sinn selbst gegeben? Betrachten wir kurz einen solchen Fall visuellen Selbsterlebens, der durch eine Erweiterung des Selbstmodells zustandekommt.

Was unterscheidet einen Rhesusaffen von einem Schimpansen? Die in unserem theoretischen Kontext interessante Antwort lautet: Ein Schimpanse kann sich selbst im Spiegel wiedererkennen, ein Rhesusaffe kann es nicht. Das bedeutet: Ein Schimpanse ist in der Lage, ein visuelles mentales Modell seines Spiegelbildes *in sein Selbstmodell einzubetten*.[41] Innerhalb einer repräsentationalen Erkenntnistheorie kann man vorsprachliche Selbsterkenntnis nämlich als die Aktivierung eines Selbstmodells oder als die Einbettung eines neuen Repräsentats in ein bereits *bestehendes* Selbst-

[41] Vgl. Metzinger 1994.

repräsentat analysieren. Wenn bereits ein Selbstmodell besteht, wie im Fall des wachen und agierenden Schimpansen, kann dieses Selbstmodell durch die Einbettung weiterer mentaler Modelle erweitert werden. Durch jede erfolgreiche Einbettung dieses Typs reichert sich der *Gehalt* des Selbstmodells an. Der Moment, in dem ein Schimpanse erkennt, daß der rote Farbtupfer, den die Versuchsleiterin seinem visuellen Gegenüber im Spiegel auf die Stirn malt, gleichzeitig mit den taktilen Empfindungen auf seiner Stirn entsteht und daß die Stirn seines visuellen Gegenübers *seine* Stirn ist, ist der Moment der *Selbsterkenntnis*: Es ist genau der Moment, indem sein Gehirn das visuelle Modell des gespiegelten Gegenübers in das bereits aktive Selbstmodell einbettet, weil es ganz bestimmte Übereinstimmungen *(relationale Homomorphien)* zwischen beiden mentalen Strukturen entdeckt hat. Nun ist der Affe sich auch subjektiv *als Spiegelbild* gegeben. Man könnte sagen, daß sein phänomenales Ich sich ausgedehnt hat. Das merkt man unter anderem manchmal daran, daß er sich nun voller Interesse sein Hinterteil im Spiegel zu betrachten beginnt, einen Aspekt seiner selbst, der ihm bisher verborgen geblieben war.

Wenn der Schimpanse sprechen könnte, würde er vielleicht auf diesen plötzlichen Wechsel seines Bewußtseinszustandes hinweisen, indem er eine restringierte Variante des „TN-Satzes" äußert und sagt: *„Ich bin dieses Spiegelbild!"* Er würde dann mit einer Identitätsaussage auf die innere Fusionierung zweier mentaler Modelle aufmerksam machen. Er erreicht dies, indem er bezugnimmt auf neue psychologische Eigenschaften und neue phänomenale Zustände seiner selbst. Selbsterkenntnis geht in solchen Fällen mit der Expansion phänomenalen Selbst*bewußtseins* einher: Was vorher das externe Bild eines anderen Wesens war, ist nun *mein* Spiegelbild. Das mentale Modell des Spiegelbildes behält zwar – zumindest in unserem eigenen Fall – den Modus der Externalität bei, wird aber nun zur Instanz einer interessanten phänomenalen Eigenschaft, der uns jetzt bereits wohlbekannten „Meinigkeit". In solchen Fällen *identifizieren wir uns* mit einem repräsentational gegebenen Teil der Welt, weil die Einbettung des entsprechenden mentalen Modells in unser Selbstmodell einhergeht mit einer Veränderung des phänomenalen Selbst. Da Subjektivität im Kern eine repräsentationale Relation ist, bedeutet bestimmte innere Zustände[42] *als seine eigenen zu erkennen* in Wirklichkeit *sie zu seinen eigenen zu machen* – also einen bestimmten Repräsentationsmodus zu wählen (Wenn man so will: durch naturalisierte Versionen der Husserlschen „Akte des Vermeinens"). Wir begegnen hier auf einer höheren Stufe wieder dem *esse est experiri* der phänomenalen Ebene – Selbsterkenntnis ist Selbsterzeugung.

Wirklich interessant sind aber solche Fälle, in denen ein informationsverarbeitendes System ein komplexes, multimodales *internes Spiegelbild* seiner selbst durch einen solchen Akt repräsentationaler Vermeinung zu

[42] Das können auch innere Zustände sein, deren repräsentationaler Gehalt im Modus phänomenaler Externalität dargestellt werden (z. B. die mentalen Modelle von Spiegelbildern).

einer Instantiierung phänomenaler „Meinigkeit“ benutzt. Ein solcher Fall tritt immer dann ein, wenn ein menschliches Subjekt den *View from Nowhere* in sich aufgebaut hat und ihn nun wieder fallen läßt.

> **P 14**: *Lassen sich das Nagelsche* ***objective self*** *und die mit seinem Auftreten verbundenen Perspektivenwechsel als Resultat mentaler Repräsentation verstehen?*

Der *View from Nowhere* ist eine mentale Operation höchster Stufe. Er setzt bestimmte repräsentationale Funktionen voraus, die vielleicht spezifisch menschlich sind und uns wahrscheinlich von allen anderen Tieren auf diesem Planeten unterscheiden. Damit man in bezug auf das Problem psychologischer Subjektivität nicht zu naiv-realistischen, essentialistischen und cartesianischen Fehlschlüssen gelangt, muß man zuallererst auf den zentralen Punkt hinweisen, der in der philosophischen Analyse des Phänomens nie verloren gehen darf: Wir haben es hier mit einem Fall von *mentaler Simulation* zu tun.

Mentale Simulationen sind Fälle, in denen ein System mentale Modelle in Abwesenheit der Standard-Kausalketten aktiviert und zu komplexen mentalen Modellen der Realität zusammenfügt. Der *Blick von nirgendwo* ist ein phänomenaler Zustand, der dadurch entsteht, daß die mentale Simulation einer *nicht-subjektzentrierten Welt* versucht wird, also die mentale Modellierung der Realität *ohne* die Einbindung eines Selbstmodells. Er ist somit auch der Versuch eines Systems, sich vorübergehend von dem zu befreien, was ich als die Benutzerfixierung seines inneren Bildes der Welt bezeichnet habe. Dieser Versuch zur Selbsttranszendenz ist natürlich ein Menschheitsthema, ein klassischer Topos nicht nur der abendländischen Philosophie. Die Mystiker aller Kulturen haben versucht, aus der partikularen Perspektive ihres endlichen Egos auszubrechen, indem sie Zustände phänomenalen Bewußtseins kultivierten, die nicht-zentriert, holistisch und frei von der Fragmentierung durch das psychologische Subjekt waren. Sie wollten in sich einen Raum schaffen, in dem sich der *Große Blick von nirgendwo* ereignen kann. Thomas Nagel spricht aber nicht von solchen Zuständen, sondern gewissermaßen von ihrer nicht-elitären Variante: Der *Blick von nirgendwo* ist eine mentale Operation, die wir *alle* beherrschen. Wir alle können vorübergehend zumindest in der Vorstellung eine objektive Perspektive auf uns selbst einnehmen.

Diese Vorstellung bleibt aber immer noch *unsere* Vorstellung. Wenn das nicht so wäre, könnten wir uns nicht an sie erinnern und auch keine philosophischen Diskussionen über ihre korrekte Interpretation im Rahmen einer Theorie des Geistes oder innerhalb einer Erkenntnistheorie führen. Vor dem Hintergrund einer naturalistischen Theorie mentaler Repräsentation ist der *View from Nowhere* mithin eine intendierte Imagination, die mentale Simulation eines nicht-zentrierten Erlebnisraums *innerhalb* eines zentrierten Erlebnisraums. In diesem mental simulierten „objektiven“ Er-

lebnisraum befindet sich ein *mentales Spiegelbild* der Person, die ihn in sich erzeugt. Dieses Spiegelbild kann zum Beispiel ein mentales Modell von Thomas Nagel sein, allein in Ozeanen von Raum und Zeit, allein unter einer unendlichen Anzahl anderer Menschen. Da es nur ein mentales Spiegelbild, aber kein Selbstmodell, innerhalb des imaginierten Bilds der Wirklichkeit gibt, ist es scheinbar auch nicht auf einen spezifischen Benutzer festgelegt. Diese Illusion entsteht genau dann, wenn wir vergessen, daß wir es sind, die diese Vorstellung absichtlich initiiert haben (zum Beispiel, wenn wir aus dem philosophischen Gedankenexperiment in einen manifesten Tagtraum übergehen). Wenn die Illusion wieder verschwindet, *entdecken* wir uns selbst im inneren Spiegel des *Blicks aus dem Nirgendwo,* und – „Heureka!" – das Simulat bricht augenblicklich zusammen, weil unser Gehirn das mentale Modell der betreffenden Person mit unserem Selbstmodell fusioniert.

Auf diese Entdeckung meiner selbst in einem sehr umfassenden inneren Spiegelbild, dem multimodalen mentalen Simulat meiner selbst als einer *fremden* Person, kann ich nun monologisch Bezug nehmen, indem ich ausrufe: *„Ich bin TM!"* Ich beziehe mich dabei nicht auf ein irreduzibles, perspektivisches Faktum, sondern auf das wiedergewonnene *Erlebnis der Perspektivität*. Genaugenommen ist diese Perspektivität niemals vollständig verloren gegangen. Ich erfahre sie jedoch in dem Moment noch einmal als besonders prägnant, in dem mein mentales Spiegelbild wieder in mein phänomenales Selbst eingebettet wird. Ähnlich wie das Umspringen der mentalen Modelle eines Neckerschen Würfels zwischen zwei Interpretationen, ist auch der Absturz aus dem *Blick von nirgendwo* in die nun wieder vollständige Perspektivengebundenheit des psychologischen Subjekts „TM" eine *reversible* Operation. Ich kann sie beliebig oft durchführen – immer begleitet von dem, was Nagel als den *philosophical thought* bezeichnet hat. Er ist das mentale Modell des immer wieder möglichen monologischen Sprechakts, der den Perspektivenwechsel begleiten kann.

Einwand: *Aber das ist ja schrecklich – die Selbstmodell-Theorie der Subjektivität psychologisiert das transzendentale Ego und erklärt das phänomenale Ich zu einer grandiosen Illusion!*

Antwort: In der Tat ist das objektive Selbst die Verdinglichung einer psychischen Fähigkeit, für die wir vielleicht bald gute wissenschaftliche Erklärungen besitzen werden. Man kann aber nicht sagen, daß das phänomenale Ich „nur" eine Illusion ist, weil diese Redeweise selbst wieder eine Perspektive und einen Getäuschten impliziert: Es gibt noch nicht einmal den Getäuschten. Es gibt in dieser Hinsicht weder Wissen noch Illusion, sondern nur ein System, das verschiedene mentale Modelle der Realität und des Selbst aktiviert, weil es bestimmte Ziele verfolgt. Außerdem ist das menschliche Selbstmodell in nicht-pathologischen Wachzuständen keinesfalls *epistemisch leer*: Es inkorporiert biologisches und metapsychologisches Wissen[43] des Systems und seiner Gattungsgeschichte, dessen enorme Komplexität

[43] Vgl. Van Gulick 1988a, 1988b.

zugleich die Grundlage seiner bisher unübertroffenen Funktionalität darstellt.

Einwand: *Wenn ich glaube, daß das phänomenale Ich letztlich nichts anderes als eine komplexe Datenstruktur in einem biologischen Gehirn ist, dann kann ich deshalb noch lange nicht glauben, daß* ***ich selbst*** *eine solche Datenstruktur bin. Woran liegt das?*

Antwort: Dadurch, daß Sie eine bestimmte philosophische Hypothese für wahr halten, ändert sich Ihr repräsentationaler Gesamtzustand – zumindest kurzfristig – nur in vernachlässigbarem Ausmaß. Wenn es richtig ist, daß „propositionale Einstellungen" (wie etwa *„Ich glaube, daß die Selbstmodell-Theorie der Subjektivität eine gute Arbeitshypothese ist"*) in das Selbstmodell eingebettete mentale Modelle von Sätzen in externen Codes sind, dann verändern sich durch eine solche Einbettung die Ihrer mentalen Selbstmodellierung zugrundeliegenden neurobiologischen *Funktionen* zunächst nicht. Dadurch, daß Sie auf der Ebene mentaler Modellierung eine bestimmte propositionale Einstellung einnehmen, indem Sie sich eine bestimmte theoretische Position *zu eigen machen*, verändert sich ihr phänomenales Modell der Realität nämlich nur geringfügig. Es mag sogar sein, daß die neurobiologischen Realisierungen der Selbstmodellierungsfunktion im Fall des Menschen nur schwach plastisch sind, weil sie in ihrer Struktur genetisch determiniert und auf diese Weise „fest verdrahtet" sind. Vielleicht fördern empirische Untersuchungen zukünftig Genaueres über die Plastizität der funktionalen Architektur zutage, die die Instantiierungsbasis für unser phänomenales Selbstbewußtsein darstellt.

Um zu *erleben*, daß das phänomenale Selbst nur ein mentales Modell – also: ein durch Eigenschaftsbindung entstandenes *repräsentationales Objekt* – ist, müßte ein System sein naiv-realistisches Selbstmißverständnis durchbrechen: Es müßte ein Modell des Selbst *als eines internen Konstrukts* aktivieren. Für ein System, das gezwungen ist, unter biologischen Imperativen zu operieren, stellt eine solche Veränderung seiner eigenen psychischen Struktur aber keinen Vorteil dar. Dafür gibt es vielleicht zwei Gründe. **Erstens** ist es repräsentational aufwendiger, den Konstruktionsprozeß mentaler Modelle metarepräsentational innerhalb dieser noch einmal darzustellen, um sie so als Modelle zu kennzeichnen. Die *Funktionalität* der betreffenden Strukturen würde dadurch jedoch nicht automatisch erhöht. Man kann sagen: Repräsentationale Selbsttransparenz stellt keinen direkten Vorteil im Kampf ums Überleben dar, würde aber einen wesentlich höheren Aufwand an Informationsverarbeitung für das System bedeuten. Vielleicht ist der naive Realismus die funktional adäquateste „epistemologische Grundannahme" für durch eine evolutionäre Wettbewerbssituation stark gebundene natürliche Repräsentationssysteme. Das heißt natürlich nicht, daß er es auch für Menschen oder künstliche Systeme sein muß. **Zweitens** birgt der Versuch, das Selbstmodell auch auf phänomenaler Ebene *als ein mentales Konstrukt* darzustellen, die Gefahr in sich, das System in einen endlosen Progreß der Selbstmodellierung zu führen und auf diese Weise zu lähmen. Biologische Systeme müssen solche Entwicklungen vermeiden oder sie sterben aus.

Vielleicht ist die hier skizzierte Theorie des Mentalen deswegen intuitiv so wenig einleuchtend und emotional so wenig ansprechend, weil das uns von der Evolution mitgegebene und unser vorphilosophisches Denken bestimmende Wirklichkeitsmodell Eigenschaftsmengen unter Ausfilterung von Ambiguitäten notorisch zu *Dingen* verobjektiviert. Dadurch werden unscharfe Datenmengen erlebnismäßig unhintergehbar verfestigt und erhalten den illusionären Charakter von Permanenz und Substantialität. Zudem ist dieses mentale Realitätsmodell durch einen inhärenten naiven Realismus gekennzeichnet, der uns auf der Ebene phänomenalen Selbstbewußtseins – dies ist vielleicht die philosophisch interessanteste Einsicht der vorangegangenen Überlegungen – in ein erlebnismäßig schwer zu transzendierendes *naiv-realistisches Selbstmißverständnis* verstrickt. Aus diesem Grund fällt es uns schwer zu begreifen, daß auch unser Ich, das phänomenale Zentrum der Welt, keine Substanz ist, sondern schlicht ein mentales Modell – wenn auch eines von allerhöchster Komplexität. Sollte es in der Zukunft einmal Menschen oder künstliche Systeme geben, die *diese Tatsache* in ihrem phänomenalen Modell der Wirklichkeit darstellen können, dann würden solche Wesen in einem völlig anderen Bewußtseinzustand leben als wir.

5.3 Mentale Selbstmodellierung und die Würde informationsverarbeitender Systeme

Ich habe im letzten Abschnitt eine Reihe von vorläufigen Antworten auf die im ersten Kapitel aufgeworfenen Fragen zu geben versucht. Diese Antworten waren Antworten aus der Perspektive einer naturalistischen Theorie des Geistes. Eine solche Theorie sieht Subjektivität als eine komplexe psychologische Eigenschaft, die von natürlich entstandenen Repräsentationssystemen instantiiert werden kann. Die vorläufigen Antworten einer solchen Theorie wurzeln in der Annahme, daß es *repräsentationale* Merkmale eines informationsverarbeitenden Systems sind, die ihm die Erzeugung eines phänomenalen Selbstbewußtseins ermöglichen, welches sowohl intentionalen als auch qualitativen Gehalt besitzt. Eine Theorie des psychologischen Erlebnissubjekts wird deshalb in ihrem Kern eine *repräsentationale Theorie* sein: Sie wird uns erklären, in welchen Fällen eine *Objektbildung* zu einer *Subjektbildung* führt. Ich werde mich nun abschließend der Frage zuwenden, welche Gestalt das Bild vom Menschen besitzt, das aus der neuen, sich in ersten schwachen Umrissen abzeichnenden, Theorie des Geistes hervortritt. Lassen sich bei der gebotenen Vorsicht in Hinblick auf mögliche zukünftige Revisionen der empirisch fundierten Theorie des Geistes, die nun langsam Form anzunehmen beginnt, anthropologische Schlüsse auf die Natur des Menschen *als eines geistigen Wesens* ziehen?

Wir sind mentale Selbstmodelle informationsverarbeitender Biosysteme und zentrieren eine kritische Teilmenge ihrer internen Repräsentate der Welt. Werden wir nicht errechnet, so gibt es uns nicht. Als geistige Wesen

sind wir somit Teil einer evolutionären Strategie: Gewisse Organismen können mit unserer Hilfe ihren internen Simulationsraum zentrieren, wesentlich differenzieren und so ein beträchtlich höheres Maß an Information verarbeiten. Dies stellte für sie einen Vorteil im Kampf ums Überleben dar. Das psychologische Subjekt ist somit ein abstraktes Organ, das von einem biologischen Organismus episodisch aktiviert wird, um gewisse Ziele zu verfolgen. Als über ihren repräsentationalen Gehalt individuierte Datenstruktur stellt es einen Teil der Eigenschaften des sie erzeugenden Systems *für* dieses intern noch einmal dar. Als *Menge von funktionalen Zuständen* – als Ensemble von Mikrozuständen also, die über die kausale Rolle betrachtet werden, die ihre jeweilige *physikalische* Instantiierungsbasis innerhalb des Systems spielt – ist es ein abstraktes Organ, durch welches das System sein inneres und äußeres Verhaltensrepertoire beträchtlich erweitert.

Mit der Erzeugung solcher komplexer funktional-repräsentationaler Zustände wird ein System zur Instanz eines neuen Sets von Eigenschaften: psychologischer Eigenschaften nämlich, die ihm traditionell auf der personalen Ebene zugeschrieben werden. Die psychologischen Eigenschaften menschlicher Wesen sind die Eigenschaften von Personen, psychologische Zustände sind Personenzustände. Die *Subjektivität* dieser Zustände besteht darin, daß sie in ein Selbstmodell eingebettet sind. Sie werden dadurch auf der Ebene inneren Erlebens für das System zu *seinen* Zuständen. Phänomenale Meinigkeit hat damit ihre Grundlage in einer Einbettungsrelation, die zwischen zwei Typen von mentalen Modellen besteht: Wir sind geistige Wesen, deren subjektive Identität aus strukturellen Merkmalen der Informationsverarbeitung innerhalb ihrer physischen Basis entsteht. Also sind wir auch in bezug auf unsere Innerlichkeit kontingent, die Produkte der Interessen des uns aktivierenden Systems und – wie dieses – endlich.

Ist dies nur eine weitere Demütigung des Menschen durch die Wissenschaft[44], die nun in der Naturalisierung des Subjekts gipfelt? Die neue Anthropologie zeigt uns als radikal sterbliche Wesen: Es gibt keinen Grund zu der Annahme, mentale Modelle könnten ohne ihre neurobiologischen Rahmenbedingungen aktiviert werden. Unsere Tragik besteht zudem darin, daß wir Wesen sind, die ihre Endlichkeit mental sehr deutlich repräsentieren können. Dadurch müssen wir in einem permanenten Konflikt mit biologischen Imperativen leben, die in den stammesgeschichtlich älteren Teilen unserer neuronalen Hardware fest verdrahtet sind. Die Sicherung unseres individuellen Überlebens ist für uns ein *struktureller* Imperativ, den wir in Standardsituationen nicht außer Kraft setzen können. Die neueren Bereiche unseres Gehirns erzeugen aber zur gleichen Zeit ein Modell des Selbst, welches uns deutlich zeigt, daß alle Versuche, sich diesem Imperativ zu fügen, letztlich zum Scheitern verdammt sind.

[44] Vgl. dazu die von Gerhard Vollmer entwickelte Taxonomie von Kränkungen des Menschen in Vollmer 1992. Besonders interessant in unserem Zusammenhang sind die Demütigungen Nr. 7 und Nr. 9 durch das Computermodell des menschlichen Geistes und die Neurobiologie des 21. Jahrhunderts.

Es mag zudem sein, daß uns eine zukünftige Theorie des Geistes wenn auch vielleicht nicht als pure Epiphänomene, so aber doch als Wesen darstellt, deren subjektive Zustände in ihrem Gehalt in viel stärkerer Weise durch bewußtseinsexterne Faktoren, also „von unten", determiniert sind, als wir *introspektiv* ahnen können. Es könnte sein, daß unsere Selbstmodelle in bezug auf die Rationalität unserer Handlungen und auf unsere kausale Rolle bei der Handlungsgenese als geistige Wesen schlicht falsch sind: Funktional adäquate innere Konfabulationen, die das phänomenale Selbst als den Initiator von Handlungen repräsentieren, während es in Wirklichkeit nur der subjektive Schatten von Aktivitäten ist, die die neuronale Maschinerie immer schon durchgeführt hat. Für solche Thesen ist es beim gegenwärtigen Stand unseres empirischen Wissens noch zu früh. Es gibt jedoch Anzeichen, die in diese Richtung deuten.[45] Als physische *Systeme* sind wir zudem Wesen, die durch einen Schleier aus tanzender Information von sich selbst und der Welt getrennt sind. Wir sind uns selbst nicht in der Direktheit eines sich selbst durchsichtigen Ichs gegeben, sondern nur durch die Produkte des von uns selbst erzeugten inneren Informationsflusses. Aufgrund von Eigenheiten der Weise, in der wir die Welt und uns selbst *für* uns selbst darstellen, wird dieser Informationsfluß nicht nur zum Medium der Wissensgewinnung sondern auch zum Schleier, der uns auf der Ebene inneren Erlebens in einem naiv-realistischen Selbstmißverständnis gefangenhält. Allein die Tatsache, daß wir diese Einsicht aussprechen und nun im Rahmen naturalistischer Theorien des Geistes mit zunehmender Genauigkeit werden formulieren können, zeigt jedoch, daß dies nicht der Endpunkt unserer Entwicklung sein muß.

Das neue Bild vom Menschen als eines endlichen Repräsentationssystems mit einer natürlichen Genese ist aber nicht nur eine Demütigung, die uns nun endgültig einer Würde beraubt, die im Grunde schon immer eine Illusion war. Ein neues Selbstverständnis ist immer auch eine Chance: Die Chance, eine neue Perspektive auf sich selbst einzunehmen und unsere rationale Selbsterkenntnis auf eine neue Ebene zu heben. Das neue Bild von uns selbst als biologischen Selbstmodellerzeugern birgt die Chance, die Merkmale, die das neue Bild als zentral zeigt, anzunehmen – und gegebenenfalls sogar zu kultivieren. Ein würdevoller Umgang mit der eigenen Natur kann somit darin bestehen, den Tatsachen ins Auge zu schauen und dann auf der Grundlage eines realistischen Selbstbildes Zielvorstellungen zu definieren. Denn auch jenseits essentialistischer Subjektmetaphysik ergeben sich eine ganze Reihe neuer Perspektiven, aus denen wir als in vieler Hinsicht wunderbare Wesen erscheinen.

Der Zustandsraum von Gehirnen ist unfaßbar groß. Die Systeme, die unsere phänomenale Realität erzeugen, besitzen einen Komplexitätsgrad, den wir *subjektiv* auch nicht ansatzweise erfassen können. Die Anzahl der ihnen möglichen Zustände ist gigantisch, zu groß um in der Periode eines einzigen Menschenlebens durchlaufen zu werden. Ich habe bereits früher

[45] Vgl. etwa Edelman 1989 und die bereits zitierten Arbeiten von Libet.

auf Paul Churchlands amüsante Milchmädchenrechnung[46] hingewiesen, derzufolge die Anzahl der kognitiven Konfigurationen, die ein menschliches Gehirn annehmen kann, ein sehr großes Vielfaches der Zahl der Elementarteilchen im gesamten Universum darstellt. Die Zahl der funktional interessanten und stabilen Zustände wird wesentlich geringer sein. Trotzdem hat uns die Theorie neuronaler Netze eine Ahnung davon vermittelt, welche Dimensionen neurobiologischer Komplexität unserem Mentalleben zugrunde liegen. Die Plastizität und der potentielle repräsentationale Reichtum unserer biologischen Basis sind enorm. Man kann vermuten, daß wir auch die Größe unseres *phänomenalen* Zustandsraums weder nutzen noch ahnen.

Unsere repräsentationalen Gesamtzustände sind sehr plastisch. In der Tat sind die Bewußtseinszustände, die durch den Informationsfluß in unseren Gehirnen entstehen, so detailreich und formbar, daß diese Tatsache häufig auch von Philosophen als Argument gegen mechanistische Versuche zur Erklärung des Geistigen im Menschen verwendet wurde. Mechanismus in der Philosophie des Geistes ist aber nicht die Sehnsucht nach der Rückkehr in ein Uhrwerkuniversum (zumindest im Bereich der psychologischen Ontologie), sondern einfach ein erkenntnistheoretisches Programm: Wir glauben, daß auch für die mentalen Eigenschaften eines Systems gilt, daß man sie allein aus seinen Teilen und den zwischen ihnen bestehenden Relationen verstehen kann. Dem mechanistischen Ansatz zufolge sind wir nicht blinde Maschinen im Sinne einer biologischen Realisierung *deterministischer, abstrakter Automaten* mit simplen ganzzahligen Übergangswahrscheinlichkeiten zwischen ihren internen Zuständen. Wir sind viel eher so etwas wie *Mengen* miteinander interagierender virtueller Maschinen, die massiv parallel arbeiten und untereinander noch einmal vernetzt sind. Das Gehirn ist ein neurobiologisch realisiertes System aus miteinander in Wechselwirkung stehenden virtuellen Maschinen, die mentale Modelle produzieren, deaktivieren, ineinander transformieren und einbetten. Aus der Selbstorganisation ihres Zusammenspiels entstehen „holistische" repräsentationale Gesamtzustände, die ein so hohes Maß an Formbarkeit und Schnelligkeit aufweisen, daß es kaum glaublich erscheinen mag, daß all dies auf der Grundlage physikalischer Prozesse geschieht: Wir müssen unser Bild davon revidieren, was in einem physikalischen Universum möglich ist.

Biologische Informationsverarbeitung ist ein weit aufregenderes Phänomen, als mythische Ätherleiber und die Metaphysik der Seele es jemals sein könnten. Die von unseren Gehirnen geöffneten Simulationsräume machen uns zu *offenen* Wesen, die eine unermeßlich große Zahl von Problemen erfassen und strategisch angehen können. Die mit der Plastizität und dem repräsentationalen Potential unserer mentalen Realitätsmodelle einhergehende Palette *phänomenaler* Zustände ist so groß, daß es keinem einzelnen menschlichen Individuum jemals gelingen wird, die Tiefe seines subjektiven Erlebnisraums vollständig auszuloten. Die *Plastizität* unseres inneren

[46] Vgl. Churchland 1989: 131f.

Realitätsmodells zeigt sich besonders prägnant in dem was Thomas Nagel als unsere *Kompetenz für Objektivität* bezeichnet hat: In unserer Fähigkeit zu *repräsentationaler Selbsttranszendenz*. Die Fähigkeit zur Abschwächung der Perspektivität unseres mentalen Modells der Realität durch den *Blick von nirgendwo* ist möglicherweise das philosophisch bedeutsamste Charakteristikum menschlicher Subjektivität. Auch wenn der *Blick von nirgendwo* am Ende doch nur eine natürlich entstandene Fähigkeit unserer Gehirne ist, die zu einer neuen Objektbildung auf einer höheren Ebene führt, kann man die Fähigkeit zur repräsentationalen Distanzierung vom Selbstmodell als unsere höchste und bedeutende kognitive Kompetenz ansehen. Vielleicht ist das Ich nicht die Grenze der Welt, sondern die Fähigkeit, diese Grenze auf immer höheren phänomenalen Ebenen neu zu ziehen.

Unser funktionales Profil ist reich, flexibel und erfolgreich. Gemessen an allen anderen informationsverarbeitenden Systemen, die wir aus der biologischen Geschichte unseres Planeten kennen oder selbst konstruiert haben, sind wir von überlegener funktionaler Komplexität. Die Breite unserer Handlungsmöglichkeiten – auch unserer *inneren* Handlungsmöglichkeiten – ist so enorm, daß sie zum Problem geworden ist. Als Gattung sind wir – bisher – äußerst erfolgreich, wir haben die biologische Evolution um eine, wenn auch wesentlich primitivere, *kulturelle* Evolution bereichert. In der Interaktion mit der Umwelt und miteinander haben wir ein beträchtliches Maß an funktionaler Flexibilität erreicht: Als Systeme sind wir nicht mehr in starre Reaktionsmuster gepreßt. Wir sind auch *ökologisch offene* Wesen, die in einer Vielzahl natürlicher oder selbstgeschaffener Umwelten existieren können. Die aktuelle technische Metapher, mit deren Hilfe wir nun für einige Zeit den Prozeß der Selbsterkenntnis vorantreiben können, ist das Netz. Wir sind Systeme, die durch Myriaden mikrofunktionaler Zustände ein großes Maß an Information gleichzeitig verarbeiten können – sogar und gerade dann, wenn diese Information „verrauscht“ oder uneindeutig ist. Auch für das phänomenale Ich, das Selbstmodell, gibt es eine neue Metapher: Da es nicht nur einen Gehalt, sondern auch ein funktionales Profil besitzt, können wir es als eine *virtuelle Maschine* interpretieren, die durch einen speziellen Mechanismus implementiert ist, der beim Menschen durch die Aktivität einer genetisch fixierten und fest verdrahteten Neuromatrix – das Körperschema – gebildet wird. Sein Fundament ist der kontinuierliche und inputunabhängige Signalaspekt des Körperpräsentats, der physische Anker des phänomenalen Ich. Dieser spezielle Mechanismus macht das Selbstmodell zu einem diskreten funktionalen Subsystem, welches in diesem Sinn eine gewisse Autonomie besitzt.

Wir bewegen uns in hochabstrakten, multidimensionalen Räumen. Die Tatsache, daß unser mentales Realitätsmodell sich durch eine starke *phänomenale Konkretheit* auszeichnet, darf nicht darüber hinwegtäuschen, daß die zugrundeliegenden Mechanismen der Informationsverarbeitung sehr komplex sind. Die Formalismen und Algorithmen, die nötig sind, um die Entstehung repräsentationalen Gehalts in unseren Gehirnen zu erfassen, sind allem Anschein nach recht abstrakt und umfangreich. Besonders die

Aktivierung eines stabilen Selbstmodells dürfte aus neurowissenschaftlicher Perspektive ein äußerst komplexes Ereignis sein, das enorme Rechenkapazitäten erfordert. Auch wenn uns die Perspektivität unseres Erlebnisraums *subjektiv* als etwas Einfaches und Gegebenes erscheint, könnte eine Vektoranalyse der zugrundeliegenden neuronalen Aktivierungszustände die entsprechenden Bewegungen des Systems als Bewegungen in sehr hochdimensionalen Räumen erweisen.

Wir sind gleichzeitig soziale Subjekte. Das heißt: Wir sind nicht nur isoliert operierende natürliche Repräsentationssysteme, sondern immer schon eingebunden in *umfassendere* Systeme der Repräsentation und Kommunikation. Menschliche Wesen, die über ihr Innenleben kommunizieren, erzeugen eine Art Interface zwischen zwei sehr verschiedenen Arten von informationsverarbeitenden Systemen, nämlich zwischen Gesellschaften und Gehirnen. Die von diesen Systemen verwendeten Codes bedienen sich verschiedener Darstellungsformen: sprachlicher Ausdrücke und mentaler Modelle. Durch Differenzen in den verwendeten Formaten ergeben sich Vermittlungsprobleme zwischen den sozial und mental generierten Modellen der Wirklichkeit und des Subjekts selbst. Trotzdem macht die Einbettung menschlicher Individuen in umfassendere Repräsentationssysteme mit einer eigenen geschichtlichen Dynamik[47] eine weitere wichtige Dimension ihres psychischen Reichtums aus.

Der Gehalt unserer Selbstmodelle und damit unseres phänomenalen Selbstbewußtseins wird durch diese *kulturelle Einbettung* stark mitbestimmt, weil wir uns mental immer auch *als* moralische Subjekte, *als* gesellschaftliche Subjekte, als Mitglieder von rationalen Sprechergemeinschaften etc. modellieren. Die phänomenale Struktur unseres subjektiven Bewußtseins dagegen wird durch interne Informationsverarbeitung determiniert und muß durch naturalistische Kategorien wie *Repräsentation, Simulation, Präsentation, Metarepräsentation, Selbstmodellierung* erläutert werden. Es sind die strukturellen Merkmale der phänomenalen Realität (zum Beispiel „Zentriertheit"), die wir analysieren müssen, wenn wir uns für das von Nagel formulierte Projekt einer objektiven Phänomenologie[48] interessieren.

Wir sind Systeme, die externe Rationalitätszuschreibungen unserer Verhaltensweisen rechtfertigen. Zumindest manchmal rechtfertigen wir solche Zuschreibungen auch dadurch, daß wir uns über alle biologischen Imperative hinwegsetzen. Das bedeutet nicht, daß alle *inneren* Abläufe kognitiver mentaler Modellierung rational rekonstruiert werden können, etwa nach dem Muster des logischen Schließens (daß man also *Realist* bezüglich der Rationalität des Mentalen sein muß). Die Aktivierung mentaler Modelle ist

[47] Man kann auch biologische Gattungen oder Gesellschaften als informationsverarbeitende Entitäten begreifen, die eine bestimmte raumzeitliche Lokalisierung besitzen und Information parallel-distribuiert verarbeiten. Dies könnten sie zum Beispiel erreichen, indem sie die „Konnektivität" zwischen einzelnen ihrer Mitglieder verändern und so über eine Sequenz von „Lernschritten" hinweg eine „Konfiguration" oder „adaptive Landschaft" in sich erzeugen.

[48] Vgl. Nagel 1981(1974): 271f.

ein assoziativer Vorgang, bei dem Ähnlichkeitsbeziehungen eine wichtigere Rolle spielen als Folgerelationen. Jedoch besitzen unsere äußeren Handlungsmuster (relativ zu gewissen Zielen) häufig eine eigentümliche Konsistenz, die es manchmal möglich macht, uns als *rationale Subjekte* zu beschreiben. Und wir *erleben* uns als rationale Subjekte, wenn wir uns selbstmental als Wesen modellieren, die Logik und Erfahrung systematisch als Erkenntnisquellen einsetzen.[49]

Man sieht: Wir sind nicht nur blinde Roboter des Genpools, die – gefangen in einem illusionären, naiv-realistischen Selbstmißverständnis – eine kurze Zeit an dessen Optimierung arbeiten und dann wieder von dem kalten, leeren Universum verschluckt werden, das sie vorübergehend erzeugt hat. Wir sind auch Systeme, die einen Grad an repräsentationaler Komplexität erreicht haben, der es ihnen erlaubt, episodisch zur Instantiierungsbasis eines neuen Sets psychologischer Eigenschaften zu werden: Wir erzeugen qualitatives Selbstbewußtsein. Dadurch, daß wir perspektivisch organisierte Repräsentationsräume in uns öffnen, bringen wir psychologische Subjektivität *in die Welt.* Aus der hier skizzierten Theorie des Geistes geht nicht hervor, daß es nicht auch einmal künstliche Systeme geben könnte, die dieselbe Art von „repräsentational fundierter Würde" besitzen, aber vom Makel der Endlichkeit, zumindest jedoch vom Makel der Kurzlebigkeit frei wären. All diese Einsichten und Spekulationen werden jedoch die Wunde nicht heilen können, die uns durch das Bewußtsein unserer radikalen Sterblichkeit geschlagen wird.

Ich habe in diesem Buch zu zeigen versucht, wie das psychologische Subjekt in seinem vollen Gehalt *in der Welt* sein kann. Denn das eigentliche Wunder ist nicht die Transzendentalität des Subjekts, sondern seine *Immanenz*: Die physikalische Welt ist wesentlich reichhaltiger, als wir zumeist gedacht haben, weil sie ein psychisches Potential besitzt. Auf ihrer Grundlage können biologische Formen der Informationsverarbeitung entstehen, die schließlich zur Aktivierung von Selbstmodellen und der Öffnung von zentrierten Simulationsräumen führen. Wenn man so will, hat sich im Menschen die Richtung der Evolution umgekehrt: Sie geht nach *innen*. Das soll heißen: Wenn *Selbstähnlichkeit* – also eine immer bessere Form der *internen* Selbstmodellierung – das Entwicklungsziel ist, dann geht es um eine repräsentational optimierte Innerlichkeit. Zur biologischen Evolution tritt dann eine psychische Weiterentwicklung hinzu. Diese psychische Evolution besteht in der Perfektion der abstrakten Organe, die ich als mentale

[49] Rationalisierende Verhaltenserklärungen müssen, im Gegensatz zu *kausalen* Verhaltenserklärungen, nicht prinzipiell in immer genauere Spezifikationen der einzelnen Erklärungsglieder überführt werden können. Wenn es stimmt, daß rationalisierende Verhaltenserklärungen auf Abstrakta Bezug nehmen, dann ist die mentale Selbstmodellierung – die unserer eigenen Vernünftigkeit *als einem auch introspektiv erlebten Phänomen* zugrundeliegt – in diesem Punkt nicht kompatibel mit der philosophischen Theorie: Sie wirft die *causal properties* und die *content properties* unserer Handlungsgründe ineins. Vgl. Bieri 1990 und Abschnitt 1.1, Fußnote 12.

Modelle bezeichnet habe. Eine besondere Bedeutung kommt dabei jenen mentalen Modellen zu, die das Potential besitzen, vorübergehend in das aktuelle Selbstmodell des Systems eingebettet zu werden.

In diesem Sinne kehrt auch eine naturalistische Theorie des Subjekts zurück zu einem klassischen Motiv der abendländischen Philosophie, das bereits in der Inschrift des apollinischen Tempels zu Delphi seinen Ausdruck gefunden hat. Eine psychologische Anthropologie, die die Fähigkeit zu komplexen und stabilen Formen mentaler Selbstmodellierung als unser wesentliches Merkmal in den Brennpunkt des theoretischen Interesses rückt, also die Erzeugung eines durch ein Selbstmodell zentrierten inneren Bewußtseinsraums als dasjenige Charakteristikum auszeichnet, das uns von allen anderen bekannten Systemen unterscheidet und ihnen überlegen macht, wird auch Antworten für die *philosophische* Anthropologie nahelegen. Wenn der Kern unseres Daseins als geistige Wesen und Subjekte phänomenaler Zustände in der Tatsache zu finden ist, daß wir zur Klasse der Selbstmodellerzeuger gehören, dann besteht unsere Würde darin, genau diese Eigenschaft anzuerkennen und zu vervollkommnen. Bei einem solchen Unternehmen spielt aber wiederum die theoretische Form der Selbsterkenntnis eine entscheidende Rolle, weil sie uns ein besseres Verständnis der biologischen, funktionalen und repräsentationalen Grundlagen der Innerlichkeit unserer subjektiven Erlebnisse ermöglicht. Das gegenwärtig im Entstehen begriffene wissenschaftlich-philosophische Bild geistiger Prozesse ist das erste Bild dieser Art in der Geschichte der Menschheit, das ein empirisches Fundament besitzt. Darum besteht bezüglich dieses Bildes auch erstmals Grund zur Hoffnung auf einen klar benennbaren Erkenntnisfortschritt. Wir befinden uns auf dem Weg zu einer neuen Theorie des Geistes.

6. Kapitel

Anhänge

6.1 Literaturverzeichnis

Das folgende Literaturverzeichnis enthält alle Werke, auf die im Text direkt Bezug genommen wurde. Besteht der Verweis im Fußnotentext aus *zwei* Jahreszahlen, so bezieht sich die erste von diesen im allgemeinen auf die Ausgabe, nach der zitiert wurde. Die zweite Jahreszahl (in Klammern) bezieht sich auf den Zeitpunkt der Erstveröffentlichung, bei neueren Texten, die bereits in deutscher Übersetzung vorliegen, häufig auf das Erscheinungsjahr der englischsprachigen Originalversion.

Abelson (1977). Siehe Schank/Abelson 1977.

Abrahamsen, A. (1991). Siehe Bechtel/Abrahamsen 1991.

Adelmann, G. (ed) (1987). *Encyclopedia of Neuroscience*, Vol. I & II. Basel, Boston, Stuttgart: Birkhäuser.

Aghajanian, G.K. *et al.* (1968). Lysergic acid diethylamide: Sensitive neuronal units in the midbrain raphe. *Science*, **161**, 706–708.

Aghajanian, G.K. *et al.* (1970). Action of psychotogenic drugs on midbrain raphe neurons. *J. Pharm. Exp. Ther.*, **171**, 178–187.

Aghajanian, G.K. *et al.* (1975). *Amine receptors in the CNS III. 5-Hydroxytryptamine in brain.* In: IVERSEN *et al.* 1975.

Albert, H. (1968). *Traktat über kritische Vernunft.* Tübingen: Siebeck.

Allport, A. (1988). *What Concept of Consciousness?* In: Marcel/Bisiach 1988.

Alvarado, C. (1989). Trends in the Study of Out-of-Body-Experiences: An Overview of Developments Since the Nineteenth Century. *Journal of Scientific Exploration*, Vol. 3, **1**, 27–42.

Anderson, J.A/Rosenfeld, E. (1988) [Eds]. *Neurocomputing – Foundations of Resarch*. Cambridge und London: MIT-Press.

Anscombe, G.E.M (1981). *Die erste Person*. In: Bieri 1981.

Anton, G. (1899). Über die Selbstwahrnehmungen der Herderkrankungen des Gehirns durch den Kranken bei Rindenblindheit und Rindentaubheit. *Archiv für Psychiatrie und Nervenkrankheiten*, **32**, 86–127.

Armstrong, D.M. (1965). *The Nature of Mind. Inaugural Lecture of the Challis Professor of Philosophy at the University of Sydney*. Auch in: Borst 1970 und Block 1980.

Armstrong, D.M. (1968). *A Materialist Theory of the Mind.* London.

Armstrong, D.M./Malcom, N. (1984)[Eds]. *Consciousness and Causality.* Oxford: Blackwell.

Baars, B. (1988). *A Cognitive Theory of Consciousness.* Cambridge: Cambridge University Press.

Barinaga, M. (1990). The Mind Revealed? *Science*, **249**, 856–858.

Bartels, A. (1984). Zur Vereinbarkeit von Kripkes Theorie der notwendigen Identität mit der Leib–Seele–Identitätsthese. *Conceptus XIX*, **47**, 39–49.

Baynes, K., Bohman, J. & McCarthy, T. (1987)[Eds]. *After Philosophy: End or Transformation?* Cambridge/MA: MIT–Press.

Bechtel, W./Abrahamsen, A. (1991). *Connectionism and the Mind – An Introduction to Parallel Processing in Networks.* Cambridge/MA: Basil Blackwell.

Beckermann, A. (1977). *Der funktionale Materialismus H. Putnams als Lösung des Problems der Vereinbarkeit intentionaler und mechanistischer Handlungserklärungen*. In: Patzig/ Scheibe/Wieland 1977.

Beckermann, A. (1977) [Hrsg.]. *Analytische Handlungstheorie*, Band 2. Frankfurt: Suhrkamp.
Beckermann, A. (1986a). *Descartes' metaphysischer Beweis für den Dualismus – Analyse und Kritik*. Freiburg: Verlag Karl Alber.
Beckermann, A. (1986b). *Kann die Künstliche Intelligenz-Forschung Fragen der Philosophie beantworten?* In: Stoyan 1986.
Beckermann, A. (1986c). Dennetts Stellung zum Funktionalismus. *Erkenntnis*, **24**, 309–341.
Beckermann, A. (1988). Sprachverstehende Maschinen – Überlegungen zu John Searles Thesen zur Künstlichen Intelligenz. *Erkenntnis*, **28**, 65–85.
Beckermann, A. (1990). Semantische Maschinen. In: *Intentionalität und Verstehen*, herausgegeben vom Forum für Philosophie Bad Homburg. Frankfurt/M.: Suhrkamp.
Beckermann, A., Flohr, H. und Kim, J. (1992)[Eds]. *Emergence or Reduction – Essays on the Prospects of Nonreductive Materialism*. Berlin/New York: Walter de Gruyter.
Benson, D.F./Greenberg, J.P. (1969). Visual Form Agnosia. *Archives of Neurology,* **20**, 82–89.
Bieri, P. (1981)[Hrsg.]. *Analytische Philosophie des Geistes.* Königstein: Hain.
Bieri, P. (1982). Nominalismus und innere Erfahrung. *Zeitschrift für philosophische Forschung*, **36**, 3–24.
Bieri, P. (1987a). *Evolution, Erkenntnis und Kognition*. In: Lütterfelds 1987.
Bieri, P. (1987b)[Hrsg.]. *Analytische Philosophie der Erkenntnis.* Frankfurt: Athenäum.
Bieri, P. (1987c). *Intentionale Systeme*: *Überlegungen zu Daniel Dennetts Theorie des Geistes.* In: Brandstätter 1987.
Bieri, P. (1989). *Schmerz: Eine Fallstudie zum Leib-Seele-Problem*. In: Pöppel 1989.
Bieri, P. (1990). *Trying Out Epiphenomenalism*. Report No. 42 der Research Group on MIND AND BRAIN, Perspectives in Theoretical Psychology and the Philosophy of Mind (ZiF), Universität Bielefeld.
Bieri, P. (1991). *Was macht Bewußtsein zu einem Rätsel?* Antrittsvorlesung an der Universität Marburg. Auch in: *Spektrum der Wissenschaft*, Oktober 1992.
Bisiach, E. (1988). *The (haunted) Brain and Consciousness.* In: Marcel/Bisiach 1988.
Bisiach, E./Vallar, G. (1988). Hemineglect in Humans. In: Boller/Grafman 1988.
Blackmore, S. (1988). *A theory of lucid dreams and OBEs.* In: Gackenbach/LaBerge 1988.
Blackmore, S. (1982a). *Beyond the Body: An Investigation of Out-of-the-Body-Experiences.* London: Granada.
Blackmore, S. (1982b). Out of body experiences, lucid dreams, and imagery: Two surveys. *Journal of the American Society for Psychical Research*, **76**, 301–317.
Blackmore, S. (1984b). A psychological Theory of the Out-of-Body-Experience. *Journal of Parapsychology*, **48**, 201–218.
Blackmore, S. (1985). Lucid Dreams and Viewpoints in Imagery: Two Studies. *Lucidity Letter*, **4**, 34–42.
Blau, U. (1985). Die Logik der Unbestimmtheiten und Paradoxien. Kurzfassung. *Erkenntnis*, **22**, 369–459.
Blau, U. (1986a). Die Paradoxie des Selbst. *Erkenntnis*, **25**, 177.
Blau, U. (1986b). Wahrheit von Innen und Aussen. *Erkenntnis*, **25**, 1–30.
Block, N. (1978). *Troubles with Functionalism*. In: Savage 1978, nachgedruckt in Block 1980.
Block, N. (1980). *Readings in Philosophy of Psychology*, Vol. 1. Cambridge/MA: Methuen.
Block, N. (1981a). *Readings in Philosophy of Psychology*, Vol. 2. Cambridge/MA: Methuen.
Block, N. (1981b). *Imagery*. Cambridge/MA: MIT-Press.
Block, N. (1983a). Mental Pictures and Cognitive Science. *Philosophical Review*, **93**, 499–542. Auch in: Lycan 1990.
Boden, M. A. (1990a). *Escaping from the Chinese Room*. In: Boden 1990b.
Boden, M. A. (1990b) [Ed]. *The Philosophy of Artificial Intelligence*. Oxford: Oxford University Press.
Bodgan, R. (1986) [Ed]. *Belief*. Oxford University Press.
Bodis-Wollner et al. (1977). Visual association cortex and vision in man: pattern–evoked occipital potentials in a blind boy. *Science,* **198**, 629.

Bodis-Wollner, I. (1977). Recovery from Cerebral Blindness: evoked potential and psychophysical measurement. *Electroencephalography and Clinical Neurophysiology*, **42**, 178.
Boër, S./Lycan, W.G. (1980). Who, Me? *Philosophical Review*, **89**.
Boër, S./Lycan, W.G. (1986). *Knowing Who.* Cambridge/MA: MIT-Press.
Bogen, J. (1969a). The other side of the brain I. Dysgraphia and dyscopia following cerebral commissurotomy. *Bulletin of the Los Angeles Neurological Societies*, **34.**
Bogen, J. (1969b). The other side of the brain II. An appositional mind. Bulletin of the Los Angeles Neurological Societies, **34**.
Boller, F./Grafman, J.(1988)[Eds]. *Handbook of Neuropsychology*; Vol. 1. New York: Elsevier.
Born, R. (1987) [Ed]. *Artificial Intelligence – The Case against.* London und Sydney: Croom Helm.
Borst, C.V. (1970) [Ed]. *The Mind/Brain Identity Theory*. London: Macmillan.
Braitenberg, V. (1984). *Vehicles – Experiments in Synthetic Psychology*. MIT Press: London und Cambridge/MA.
Brandstätter, J. (1987)[Hrsg.]. *Struktur und Erfahrung in der psychologischen Forschung.* Berlin: de Gruyter.
Brewer, W. (1987). *Schemas versus mental models in human memory.* in: McGinn 1987.
Bruce, B./Brewer, W. (1980). *Theoretical Issues in Reading Comprehension.* Hillsdale/NJ.
Brunner, O., Conze, W., Koseleck, R. (1984)[Hrsg.]. *Geschichtliche Grundbegriffe*, Band 5. Stuttgart: Klett–Cotta.
Bubner, R. *et al.* (1970)[Hrsg.]. *Hermeneutik und Dialektik*, I. Tübingen.
Bunge, M. (1980). *The Mind-Body Problem*. Oxford: Pergamon.
Buschlinger, W. (1993). *Denk-Kapriolen? Gedankenexperimente in Naturwissenschaft, Ethik und Philosophy of Mind.* Würzburg: Königshausen und Neumann.
Butler, R. [Ed](1962). *Analytical Philosophy First Series.* Oxford.
Butterfield, J.F. (1968). Subjective (induced) color television. *Society of Motion Picture and Television Engineers Journal*, **79**, 1025–1028.
Butterfield, J.F. (1970). Subjective color created by black-and-white animation. *Society of Motion Picture and Television Engineers Journal*, **79**, 523–526.

Calvin, W. (1989a). *The Cerebral Symphony: Seashore Reflections on the Structure of Consciousness.* New York: Bantam.
Calvin, W. (1989b). A Global Brain Theory. *Science*, **240**, 1802f.
Campbell, K. (1970). *Body and Mind.* New York.
Casey, Joan F. (1992). *Ich bin Viele – Eine ungewöhnliche Heilungsgeschichte.* Reinbek bei Hamburg: Rowohlt.
Castañeda, H.N. (1982). *Sprache und Erfahrung.* Frankfurt: Suhrkamp.
Chase, M. und Weitzmann, E.(1983)[Eds]. *Sleep disorders: Basic and clinical research.* New York.
Chase, W.G. (1973)[Ed]. *Visual information processing.* New York: Academic Press.
Churchland, P.M. (1970). The Logical Character of Action Explanations. *Philosophical Review*, **79** (2), 214–236.
Churchland, P.M. (1979). *Scientific Realism and the Plasticity of Mind.* Cambridge: Cambridge University Press.
Churchland, P.M. (1981a). Eliminative Materialism and the Propositional Attitudes. *Journal of Philosophy*, **78** (2), 67–90.
Churchland, P.M. (1981b). Functionalism, Qualia, and Intentionality. *Philosophical Topics*, **12**, 121–145.
Churchland, P.M. (1984). *Matter and Consciousness.* Cambridge: MIT Press.
Churchland, P.M. (1985b). Reduction, Qualia, and the Direct Introspection of Brain States. *Journal of Philosophy*, **82**, 8–28.
Churchland, P.M. (1988a). Folk Psychology and the Explanation of Human Behavior. *Proceedings* of the Aristotelian Society, **suppl. 62**, 209–221.
Churchland, P.M. (1988c). *Reduction and the neurobiological Basis of Consciousness.* In: Marcel/Bisiach 1988.

Churchland, P.M. (1989). *A Neurocomputational Perspective*. Cambridge/MA und London: MIT–Press.
Churchland, P.M. (1998). Conceptual similarity across sensory and neural divesity: The Fodor/Lepore challenge answered. *Journal of Philosophy,* **65,** 5–32.
Churchland, P.S. (1986). *Neurophilosophy*: *Toward a Unified Understanding of the Mind-–Brain*. Cambridge: MIT Press.
Churchland, P.S/Sejnowski, T.J. (1990). *Neural Representation and Neural Computation*. In: Lycan 1990.
Churchland, P.S/Sejnowski, T.. (1992). *The Computational Brain.* Cambridge/MA, London: MIT–Press. Deutsch: *Grundlagen zur Neuroinformatik und Neurobiologie.* Braunschweig/ Wiesbaden: Vieweg.
Clark, A. (1989). *Microcognition – Philosophy, Cognitive Science, and Parallel Distributed Processing.* Cambridge/MA: MIT Press.
Cooper, L.A. (1976). Demonstration of a mental analog of an external rotation. *Perception and Psychophysics*, **19**, 298–302.
Cooper, L.A./Shephard, R.N. (1973). *Chronometric studies of the rotation of mental images.* In: Chase 1973.
Cowey, A. (1979). Cortical Maps and Visual Perception. *Quarterly Journal of Experimental Psychology*, **31**, 1.
Cowey, A./Stoerig, P. (1991b). *Reflections on Blindsight.* In: Milner/Rugg 1991.
Craik, K.J.W. (1943). *The Nature of Explanation*. Cambridge University Press.
Cramer, K. *et al* (1990). *Theorie der Subjektivität*. Frankfurt: Suhrkamp.
Crane, H./Piantanida, T.P. (1983). On seeing reddish green and yellowish blue. *Science*, **221**, 1078–180.
Crick, F./Mitchison, G. (1983). The function of dream sleep. *Nature*, **304**, 111–114.
Crick, F. (1984). The function of the thalamic reticular complex: the searchlight hypothesis. *Proceedings of the National Academy of Science*, **81**, 4586–4590.
Crick, F./Koch, C. (1990a). Towards a neurobiological Theory of Consciousness. *Seminars in the Neurosciences*, **2**, 263–275.
Crick, F./Koch, C. (1990b). Some reflections on visual awareness. *Symposia on Quantitative Biology, The Brain,* **55**, 953–962. Plainview/NY: Cold Spring Harbor Press.
Cummins, R. (1983). *The nature of psychological explanation*. Cambridge/MA: MIT-Press.
Cummins, R. (1989). *Meaning and mental representation*. Cambridge/MA: MIT-Press.
Cytowic, R.E. (1989). *Synesthesia – A Union of the Senses.* Springer: New York.

Damasio, A. (1987). *Agnosia*. In: Adelmann 1987, Vol. I.
Damasio, A.R. *et al.* (1982). Prosopagnosia: Anatomic basis an behavioral mechanisms. *Neurology*, **32**, 331–341.
Davidson, D. (1981). *Mentale Ereignisse*, in Bieri 1981. Im Original(1970): *Mental Events*; u.a. in Davidson, D. (1980). *Essays on Actions and Events*, New York.
Davidson, D. (1991). *Thinking Causes.* [Mit zwei Kommentaren von Jaegwon Kim und Ernest Sosa] Report No. 67 der Research Group on MIND AND BRAIN, Perspectives in Theoretical Psychology and the Philosophy of Mind (ZiF), Universität Bielefeld. Auch in: Heil/Mele 1993.
Davidson, D./Harman, G. (1972)[Eds]. *Semantics of Natural Language*. Ithaca.
Davis, L. (1982). Functionalism and absent Qualia. *Philosophical Studies*, **41**, 121–145.
Davis, W. (1985). *Passage of Darkness: The Ethnobotanology of the Haitian Zombie*. Chapel Hill and London: University of North Carolina Press.
Dennett, D.C. (1969). *Content and Consciousness.* London: Routledge and Kegan Paul.
Dennett, D.C. (1971). Intentional Systems. *Journal of Philosophy*, **8**, 87–106. Nachgedruckt in Dennett 1978a, deutsche Übersetzung in Bieri 1981.
Dennett, D.C. (1976a). *Conditions of Personhood.* In Rorty, A. (Ed) 1976; nachgedruckt in Dennett 1978a, deutsche Übersetzung in Bieri 1981.
Dennett, D.C. (1976b). Are Dreams Experiences? *Philosophical Review*, **LXXIII**, 151–171. Nachgedruckt in Dennett 1978a und Dunlop 1977.

Dennett, D.C. (1978a). *Brainstorms: Philosophical Essays on Mind and Psychology*. Montgomery/VT: Bradford Books.
Dennett, D.C. (1979). *On the Absence of Phenomenology*, in Gustafson/Tapscott 1979: 93–114.
Dennett, D.C. (1981). Wondering where the yellow went. *The Monist*, **64**, 102–108.
Dennett, D.C. (1987b). *The Intentional Stance*. Cambridge/MA und London: MIT Press.
Dennett, D.C. (1988). *Quining Qualia*. In Marcel/Bisiach 1988: 42–77.
Dennett, D.C. (1991). *Consciousness Explained*. Boston/Toronto/London: Little, Brown and Company.
Dennett, D.C./Hofstadter, H. (1981). *Einsicht ins Ich – Fantasien und Reflexionen über Selbst und Seele*. Stuttgart: Klett–Cotta.
Dennett, D.C./Humphrey, N. (1989). Speaking for Ourselves: An Assessment of Multiple Personality Disorder. *Raritan: A Quarterly Review*, **Vol. IX**, No. 1, 68–98.
Dentan, R.K. (1988). *Lucidity, Sex, and Horror in Senoi-Dreamwork*. In: Gackenbach & Laberge 1988.
Descartes, R. 1972 (1641). *Meditationen über die Grundlagen der Philosophie mit den sämtlichen Einwänden und Erwiderungen*. Hamburg: Meiner.
Dijk, W./Kintsch, A. (1983). Toward a model of text comprehension and production. *Psychological Review,* **85**, 363 – 94.
Dittrich, A. (1985). *Ätiologie–unabhängige Strukturen veränderter Wachbewußtseinszustände*. Stuttgart: Enke.
Dretske, F. (1981). *Knowledge and the Flow of Information*. Cambridge/MA: MIT–Press.
Dretske, F. (1986). *Misrepresentation*. In: Bodgan 1986.
Dretske, F. (1988). *Explaining Behavior – Reasons in a World of Causes*. Cambridge/MA und London: MIT-Press.
Dreyfus, H.L. (1985). *Die Grenzen künstlicher Intelligenz. Was Computer nicht können*. Königstein.
Dreyfus, H.L./Dreyfus, S.E. (1987). *Künstliche Intelligenz – von den Grenzen der Denkmaschine und dem Wert der Intuition*. Reinbek bei Hamburg: Rowohlt.
DSM-III-R (1989). *Diagnostische Kriterien und Differentialdiagnosen des diagnostischen und statistischen Manuals psychischer Störungen DSM-III-R*. Weinheim, Basel: Beltz. [Titel der von der American Psychiatric Association herausgegebenen Originalausgabe: *Diagnostic and Statistical Manual of Mental Disorders*, Third Edition, Revised. (1987) Washington/DC: American Psychiatric Association].
Dunlop, C. (1977)[Ed]. *Philosophical Essays on Dreaming*. Cornell University Press.

Eccles, J. (1976). *Brain and Free Will*. In: Globus/Maxwell/Savodnik 1976.
Eccles, J. (1982). *Das Ich und sein Gehirn*. In: Popper/Eccles 1982.
Eccles, J.C. (1966)[Ed]. *Brain and conscious Experience*. New York: Springer.
Edelman, G.E. (1987). *Neural Darwinism – The Theory of Neuronal Group Selection*. New York: Basic Books.
Edelman, G.M. (1989). *The Remembered Present – A Biological Theory of Consciousness*. New York: Basic Books.
Eimer, M. (1990). *Informationsverarbeitung und mentale Repräsentation*. Berlin: Springer.
Engel, A. *et al.* (1991a). Interhemispheric Synchronization of Oscillatory Neuronal Responses in Cat Visual Cortex. *Science*, **252**, 1177–1179.
Engel, A. *et al.* (1991c). Synchronization of oscillatory neuronal responses between striate and extrastriate visual cortical areas of the cat. *Proceedings of the National Academy of Science*, Volume 88, 6048–6052.
Engel, A. *et al.* (1992a). Temporal coding in the visual cortex: new vistas on integration in the nervous system. *Trends in Neurosciences*, Volume 15, **6**, 218–226.
Engel, A./König, P./Schillen, T. B. (1992b). Why does the cortex oscillate? *Current Biology*, Volume 2, **6**, 332–334.
Engel, A./König, P./Singer, W. (1991b). Direct physiological evidence for scene segmentation by temporal coding. *Proceedings of the National Academy of Science*, Volume 88, 9136–9140.

Engel, A./König, P./Singer, W. (1992c). *Correlated Neuronal Firing: A Clue to the Integrative Functions of Cortex?* In: Taylor *et al.* 1992: 125–139.
Enzensberger, H.M. (1967)[Ed]. *Kursbuch*, **8** (März 1967). Frankfurt.
Eysenck, M.W./Keane, M.T. (1990). *Cognitive Psychology – A Students Handbook*. London/Hove/Hillsdale: Erlbaum.

Farrell, B. (1950). Experience. *Mind*, **50**, 170–198.
Feigl, H. (1960). *Mind-body, not a pseudo-problem.* In: Hook 1960, auch in Borst 1970.
Feyerabend, P. (1981). *Mentale Ereignisse und das Gehirn*. In: Bieri 1981.
Finke, R.A. (1989). *Principles of mental Imagery*. Cambridge/MA und London: MIT-Press.
Flanagan, O. (1991). *The Science of the Mind* (2nd Edition). Cambridge/MA: MIT-Press.
Flohr, H. (1991). Brain processes and phenomenal consciousness. A new and specific hypothesis. *Theory and Psychology*, **1**, 245–262.
Flohr, H. (1992a). *Qualia and brain processes.* In: Beckermann/Flohr/Kim 1992.
Flohr, H. (1992b). Die physiologischen Bedingungen des phänomenalen Bewußtseins. *Forum für interdisziplinäre Forschung*, **1**, 49–55.
Fodor, J./Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, **28**, 3–71.
Fodor, J.A. (1972); siehe Block/Fodor 1972.
Fodor, J.A. (1975). *The Language of Thought.* Cambridge/MA: Harvard University Press.
Fodor, J.A. (1981). *Representations.* Cambridge/MA: MIT-Press.
Fodor, J.A. (1983). *The Modularity of Mind.* Cambridge/MA: MIT-Press.
Fodor, J.A. (1984). Semantics, Wisconsin Style. *Synthese*, **59**, 231–250.
Ford, K.M./Hayes, P.J.(1991)[Eds]. *Reasoning Agents in a Dynamic World.* Greenwich, CT: JAI Press.
Foss, J. (1987). On the Logic of What It Is Like To Be a conscious Subject. *Australasian Journal of Philosophy*, **67**, 205–220.
Frank, M. (1991). *Selbstbewußtsein und Selbsterkenntnis.* Stuttgart: Reclam.
Frankfurt, H.G. (1981). *Willensfreiheit und der Begriff der Person*. In: Bieri 1981.
Freksa, C./Habel, C. (1990)[Hrsg.]. *Repräsentation und Verarbeitung räumlichen Wissens.* Informatik Fachberichte 245, Subreihe Künstliche Intelligenz. Berlin: Springer.
Freud, S. (1947). *Eine Schwierigkeit der Psychoanalyse*. Gesammelte Werke, Band XII (Werke aus den Jahren 1917–1920). Frankfurt: Fischer.

Gackenbach, J. (1978). *A Personality and Cognitive Style Analysis of Lucid Dreaming*. Dissertation Abstracts International, 39, 3487B; University Microfilms No. 79–01560.
Gackenbach, J., Curren, R., & Cutler,G. (1983). Presleep determinants and postsleep results of lucid versus vivid dreams. *Lucidity Letter*, **2** (2).
Gackenbach, J./LaBerge, S. (1988)[Eds]. *Conscious Mind, Sleeping Brain*. New York/London: Plenum Press.
Garfield, P. (1975). Psychological concomitants of the lucid dream state. *Sleep Research*, **4**, 183.
Gazzaniga, M.S. (1988). *Brain Modularity: Towards a Philosophy of Conscious Experience*. In: Marcel/Bisiach 1988: 218–239.
Gazzaniga, M.S./LeDoux, J.E. (1978). *The Integrated Mind.* New York: Plenum Press.
Gebauer, M. (1992). *The View from Nowhere – Nachbemerkungen des Übersetzers anstelle eines Nachworts.* Unveröffentlichtes Manuskript.
Gentner, D./Stevens, A.L. (1983). *Mental Models.* Hillsdale/NJ: Erlbaum.
Ginet, C./Shoemaker, S. (1983)[Eds]. *Knowledge and Mind.* Oxford University Press.
Globus, G.G. (1976)[Eds]. *Consciousness and the Brain.* London/New York: Plenum Press.
Gordon, G., Maxwell, G. und Savodnik, I. (1976)[Eds]. *Consciousness and the Brain. A scientific and philosophical Inquiry*. New York/London: Plenum Press.
Goschke, T./ Koppelberg, D. (1990). Connectionist representation, semantic compositionality, and the instability of concept structure. *Psychological Research*, **52**, 253–270.
Goschke, T./Koppelberg, D. (1991). The concept of representation and the representation of concepts in connectionist models. In: Ramsey/Rumelhart/Stich 1991.

Green, C. (1968). *Lucid Dreams*. Oxford.
Gregory, R. (1986). *Odd Perceptions*. Methuen: Andover, Hants.
Gregory, R. (1988). *Consciousness* and con-science. In: Marcel/Bisiach 1988.
Griffiths, P. (1986). *On Being Mindless: Buddhist Meditation and the Mind-Body Problem*. La Salle: Open Court.
Gulyás, B./Oltoson, D./Roland, P. (1993) [Eds]. *Functional Anatomy of the Human Visual Cortex*. Oxford: Pergamon Press.
Gustafson, D.F./Tapscott, B.L. (1979)[Eds]. *Body, Mind, and Method*. Dordrecht: D. Reidel.

Hardin, C.L. (1988). *Color for Philosophers*. Indianapolis/Cambridge: Hackett.
Hardin, C.L. (1990). *Color and Illusion*. In: Lycan 1990.
Hastedt, H. (1988). *Das Leib–Seele–Problem*. Frankfurt: Suhrkamp.
Haugeland, J. (1982). Weak Supervenience. *American Philosophical Quarterly*, **19** (1), 93–103.
Haugeland, J. (1987). *Künstliche Intelligenz – Programmierte Vernunft?* Hamburg: McGraw-Hill.
Hebb, D. (1949). *The Organisation of Behavior*. New York: Wiley.
Heil, J./Mele, A. (1993)[Eds]. *Mental Causation*. Oxford: Clarendon Press.
Heimann, (1989). *Zerfall des Bewußtseins in der Psychose*. In: Pöppel 1989.
Helm, G. (1991). *Symbolische und konnektionistische Modelle der menschlichen Informationsverarbeitung*. Berlin: Springer. (Informatik-Fachberichte, **288**).
Henrich, D. (1967). *Fichtes ursprüngliche Einsicht*. Frankfurt.
Henrich, D. (1970). *Selbstbewußtsein: kritische Einleitung in eine Theorie*. In: Bubner *et al.* 1970.
Herken, R. (1988)[Ed]. *The Universal Turing Machine – A Half-Century Survey*. Hamburg/Berlin: Kammerer & Unverzagt.
Herrmann, T. (1988). *Mentale Repräsentation – ein erklärungsbedürftiger Begriff*. Arbeiten der Forschungsgruppe Kognition am Lehrstuhl III der Universität Mannheim, Bericht Nr. **42**.
Hill, C.S. (1981). Why Cartesian Intuitions are compatible with the Identity Thesis. *Philosophy and Phenomenological Research*, **42**.
Hinton, G./Sejnowski, T. (1986). *Learning and relearning in Boltzmann Machines*, in: Rumelhart/McClelland 1986.
Hobson, A. (1987). *Dreaming*, in: Adelmann 1987, Vol. I, 338–340.
Hobson, J.A./McCarley R.W. (1977). The brain as a dream–state generator: An activation–synthesis hypothesis of the dream process. *American Journal of Psychiatry*, **134**, 1335–1348.
Hook, S. (1960)[Ed]. *Dimensions of Mind*. New York: Macmillan/Collier Books.
Hopfield, J.J. (1982). Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *PNAS USA*, **79**, 2554–2558.
Horgan, T. (1983)[Ed]. Spindel Conference: Supervenience. In: *The Southern Journal of Philosophy*, Vol. XXII, Supplement.
Horstmann, R.-P. (1990). *Gibt es ein philosophisches Problem des Selbstbewußtseins?* In: Cramer/Fulda/Horstmann/Pothast 1990.
Humphrey, N. (1989). Siehe: Dennett/Humphrey 1989.
Humphrey, N. (1992). *A History of the Mind*. New York: Simon & Schuster.
Humphreys, G./Davies, M. (im Druck)[Eds]. *Approaches to Consciousness*. Oxford: Basil Blackwell.
Hurvich, L.M. (1981). *Color Vision*. Sunderland/MA: Sinauer.

Iggo, A. (1973). *Handbook of Sensory Physiology*, Vol. II. Berlin, Heidelberg, New York.
Irwin, H. (1985). *Flight of Mind*. Metuchen/NJ & London: Scarecrow Press.
Iversen, L.L. *et al.* (1975). *Handbook of Psychopharmacology,* Vol. 6. Plenum Press.

Jackendoff, R. (1987). *Consciousness and the Computational Mind*. Cambridge/MA, London: MIT-Press.
Jackson, F. (1982). Epiphenomenal Qualia. *Philosophical Quarterly*, **32**, 127–136.

Jackson, F. (1986). What Mary didn't know. *Journal of Philosophy*, **83**, 291–295.
Jacobs, B.L./Trulson, M.E. (1979). Mechanism of action of LSD. *American Scientist*, **67**, 396–404.
Johnson-Laird, P.N. (1983). *Mental models: towards a cognitive science of language, inference and consciousness.* Cambridge University Press.
Johnson-Laird, P.N. (1988). *A computational analysis of consciousness.* In: Marcel/Bisiach 1988.
Johnson-Laird, P.N. (1989). *Mental Models.* In: Posner 1989.

Katz, J./Melzack, R. (1990). Pain „Memories" in Phantom Limbs: Review and Clinical Observations. *Pain*, **43** (3), 319–336.
Kemmerling, A. (1991). Mentale Repräsentationen. *Kognitionswissenschaft*, **1**, 47–57.
Kennedy, M.B. (1989). Regulation of neuronal function by calcium. *Trends in Neurosciences*, **12**, 417–420.
Kiefer, E. (1988). *Metakognition, Introspektion und verbale Berichte über sich selbst.* Unveröffentlichte Dissertation.
Kim, J. (1978). Supervenience and Nomological Incommensurables. *American Philosophical Quarterly*, **15**, 149–156.
Kim, J. (1979). Causality, Identity, and Supervenience in the Mind-Body Problem. *Midwest Studies in Philosophy*, **4**, 31–49.
Kim, J. (1982). Psychophysical Supervenience as a Mind-Body Theory. *Cognition and Brain Theory*, **5**, 129–147.
Kim, J. (1982). Psychophysical Supervenience. *Philosophical Studies*, **41**, 51–70.
Kim, J. (1984). Concepts of Supervenience. *Philosophy and Phenomenological Research*, **45**, 153–176.
Kim, J. (1984). Epiphenomenal and Supervenient Causation. *Midwest Studies in Philosophy*, **9**, 257–270.
Kim, J. (1984). Supervenience and Supervenient Causation. *Southern Journal of Philosophy*, **22**, Spindel Conference Supplement on Supervenience, 45–56.
Kim, J. (1985). Supervenience, Determination and Reduction. *Journal of Philosophy*, **82**, 616–618.
Kim, J. (1987). „Strong" and „Global" Supervenience Revisited. *Philosophy and Phenomenological Research*, Vol. XL*VIII,* **2**, 315–326.
Kim, J. (1989). The Myth of Nonreductive Materialism. *Proceedings and Adresses of The American Philosophical Association*, **63** (3), 31–47.
Kim, J. (1990). *Explanatory Exclusion and the Problem of Mental Causation.* In: Villanueva (Ed) 1990.
Kinsbourne, M. (1988). *Integrated Field Theory of Consciousness.* In: Marcel/Bisiach 1988.
Kirk, R. (1974). Sentience and Behavior. *Mind*, **83**, 43–60.
Kitcher, P. (1984). In Defense of intentional Psychology. *Journal of Philosophy*, **81**, 89–106.
Kosslyn, S.M. (1980). *Image and Mind.* Cambridge/MA: Harvard University Press.
Kosslyn, S.M. (1981). The medium and the message in mental imagery: A theory. *Psychological Review*, **88**, 46.
Kosslyn, S.M. (1983). *Ghosts in the Mind's Machine.* New York: Norton.
Kosslyn, S.M./Shwartz, S.P. (1977). A simulation of visual imagery. *Cognitive Science*, **1**, 265–295.
Krämer, S. (1993)[Hrsg.]. *Geist, Gehirn, Künstliche Intelligenz – Zeitgenössische Modelle des Denkens.* Berlin, New York: de Gruyter.
Kreiter, A. K./Singer, W. (1992). Oscillatory Neuronal Responses in the Visual Cortex of the Awake Macaque Monkey. *European Journal of Neuroscience*, Volume 4, 369–375.
Kripke, S.A. (1971). *Identity and Necessity.* In: Munitz 1971.
Kripke, S.A. (1972). *Naming and Necessity.* In: Davidson/Harman 1972.
Kukla, R. (1992). Cognitive Models and Representation. *British Journal of Philosophy of Science*, **43**, 219–232.
Kurthen, M. (1990). *Das Problem des Bewußtseins in der Kognitionswissenschaft – Perspektiven einer „Kognitiven Neurowissenschaft".* Stuttgart: Enke.

LaBerge, S. (1980). Induction of lucid dreams. *Sleep Research*, **9**, 138.
LaBerge, S. (1980). *Lucid Dreaming: An exploratory study of consciousness during sleep.* Dissertation Abstracts International, **41**, 1966B; University Microfilms No. 80–24691.
LaBerge, S. (1980). Lucid dreaming as a learnable skill: A case study. *Perceptual and Motor Skills*, **51**, 1039–1041.
LaBerge, S. (1988). *Lucid dreaming in Western literature.* In: Gackenbach/Laberge 1988.
LaBerge, S. (1988). *The psychophysiology of lucid dreaming.* In: Gackenbach/Laberge 1988.
LaBerge, S. *et al.* (1981). Psychophysiological correlates of the initiation of lucid dreaming. *Sleep Research*, **10**, 149.
Lackner, J. R. (1988). Some proprioceptive Influences on the Perceptual Representation of Body Shape and Orientation. *Brain*, **111**, 281–297.
Langton, C.G. (1989). *Artificial Life.* Redwood City: Addison-Wesley.
Lanz, P. (1989). *Vom Geist zum Gehirn – Ein Themenwechsel?* In: Pöppel 1989.
LeDoux, J.E., Wilson, D.H. und Gazzaniga, M.S. (1977). A Divided Mind: Observations of the Conscious Properties of the Separated Hemispheres. *Annals of Neurology*, **2**, 417–421.
Leplin, J. (1979). Theoretical Explanation and the Mind–Body Problem. *Philosophia*, **8**, 672–688.
Leuner, H.-C. (1981). *Halluzinogene.* Bern, Stuttgart, Wien: Hans Huber.
Levin, J. (1986). Could Love be Like a Heatwave? Physicalism and the Subjective Character of Experience. *Philosophical Studies*, **49**(2), 245–261; auch in: Lycan 1990.
Levin, M. (1975). Kripke's Argument against the identity thesis. *Journal of Philosophy*, **72**, 149–167.
Levin, M. (1979). *Metaphysics and the Mind-Body Problem*. Oxford University Press.
Lewis, D. (1966). An Argument for the Identity Theory. *Journal of Philosophy*, **63**. Deutsch in Lewis 1989 und Beckermann 1977.
Lewis, D. (1971). Analog and Digital. *Nous*, **5**, 321–327.
Lewis, D. (1972). Psychophysical and Theoretical Identifications. *Australasian Journal of Philosophy*, **50**, 249–258. Deutsch in Lewis 1989.
Lewis, D. (1981). *Mad Pain and Martian Pain*. In: Block 1980, wiederabgedruckt und um ein Postskript erweitert in Lewis 1983, deutsch in Lewis 1989.
Lewis, D. (1983). *Philosophical Papers*, Vol I. Oxford Unversity Press.
Lewis, D. (1988). What Experience Teaches. *Proceedings of the Russelian Society*. University of Sydney; auch in: Lycan 1990.
Lewis, D. (1989). *Die Identität von Körper und Geist*. Frankfurt: Klostermann.
Libet, B. (1973). *Electrical stimulation of cortex in human subjects and conscious memory aspects.* In: Iggo 1973.
Libet, B. (1981). The experimental Evidence for subjective Referral of a sensory Experience backwards in Time: Reply to P.S. Churchland. *Philosophy of Science*, **48**, 182–197.
Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*, **8**, 529–566.
Libet, B. *et al.* (1979). Subjective Referral of the Timing for a conscious sensory Experience: A functional Role for the somatosensory specific Projection System. *Brain*, **102**, 193–224.
Linke, D.B./Kurthen, M. (1988). *Parallelität von Gehirn und Seele.* Stuttgart: Enke.
Locke, D. (1976). Zombies, Schizophrenics and purely physical Objects. *Mind*, **85**, 97–99.
Lütterfelds, W. (1987)[Hrsg.]. *Transzendentale oder Evolutionäre Erkenntnistheorie?* Darmstadt: Wissenschaftliche Buchgesellschaft.
Lycan, W.G. (1973). Inverted Spectrum. *Ratio*, **15**.
Lycan, W.G. (1979). A new Lilliputian Argument against Machine Functionalism. *Philosophical Studies*, **35**.
Lycan, W.G. (1987). *Consciousness.* Cambridge/MA, London: MIT-Press.
Lycan, W.G. (1989). *Ideas of Representation*. In: Weissbord 1989.
Lycan, W.G. (1990)[Ed]. *Mind and Cognition*. Cambridge/MA, Oxford: Basil Blackwell.

Mackay, D.M. (1966). *Cerebral organisation and the conscious control of action*. In: Eccles 1966.
Malcolm, N. (1959). *Dreaming*. London & New York.
Malcolm, N. (1988). Subjectivity. *Philosophy*, **63**, 147–160.
Maloney, J.C. (1985). About being a Bat. *Australasian Journal of Philosophy*, **63**(1), 26–49.
Mandl, H./Spada, H. (1988)[Hrsg.]. *Wissenspsychologie*. München/Weinheim: Psychologie Verlags Union.
Mandler, G. (1985). *Cognitive Psychology: An Essay in Cognitive Science*. Hillsdale/NJ: Erlbaum.
Marcel, A. (1983a). Conscious and unconscious perception: experiments on visual masking and word recognition. *Cognitive Psychology*, **15**, 197–237.
Marcel, A. (1983b). Conscious and unconscious perception: an approach to the relations between phenomenal experience and perceptual processes. *Cognitive Psychology*, **15**, 238–300.
Marcel, A. (1988). *Phenomenal Experience and Functionalism*. In: Marcel/Bisiach 1988.
Marcel, A./Bisiach, E. (1988)[Eds]. *Consciousness in contemporary Science*. Oxford: Clarendon Press.
Marr, D.(1982). *Vision – A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: W.H. Freeman and Company.
Marshall, J.C./Morton, J. (1978). *On the mechanics of Emma*. In: Sinclair/Jarvella/Levelt 1978.
McCarley, R. (1983). *REM dreams, REM sleep, and their isomorphisms*. In: Chase/Weitzmann 1983
McCarley, R./Hobson, J. (1979). *The form of dreams and the biology of sleep*. In: Wolman 1979.
McClelland, J.L., Rumelhart, D.E.& The PDP Research Group (1986)[Eds]. *Parallel distributed processing. Explorations in the microstructure of cognition*. Volume 2. Cambridge/MA: MIT Press.
McGinn, C. (1980). *Anomalous Monism and Kripke's Cartesian Intuitions*. In: Block 1980.
McGinn, C. (1983). *The subjective View*. Oxford: Oxford University Press.
McGinn, C. (1987). Critical Notice on *The View from Nowhere* by Thomas Nagel. *Mind*, 263–272.
McGinn, C. (1989). *Mental Content*. Oxford.
McGinn, C. (1991). *The Problem of Consciousness: Essays toward a Resolution*. Oxford: Basil Blackwell.
McMullen, C. (1985). „Knowing what it's like" and the essential indexical. *Philosophical Studies*, **48**, 211–234.
McMurrin, S.M. (1980)[Ed]. *The Tanner Lectures on Human Values – 1980*. Salt Lake City: University of Utah Press and Cambridge University Press.
Mead, G.H. (1964). The social self. In: *Selected Writings of George Herbert Mead* [Ed: Reck, A.J.]. Indianapolis/IN: Bobbs-Merrill.
Melzack, R. (1989). Phantom Limbs, the Self and the Brain: The D.O. Hebb Memorial Lecture. *Canadian Psychology*, **30** (1), 1–16.
Melzack, R. (1992). Phantom Limbs. *Scientific American*, **266** (4); 90–96.
Melzack, R./Israel, R./Lacroix, R./Schultz, G. (1997). Phantom limbs in people with congenital limb deficiency or amputation in early childhood. *Brain,* **120** (Pt 9), 1603–20.
Mesulam, M-M. (1987). Neglect (Selective Inattention). In: Adelmann 1987, Vol. II.
Metzinger, T. (1985). *Neuere Beiträge zur Diskussion des Leib-Seele-Problems*. Frankfurt/Bern/New York: Peter Lang.
Metzinger, T. (1989). Zentriertes Bewußtsein – Die Selbstmodell-Theorie der Subjektivität. *Berichte des XIII. Internationalen Wittgenstein-Symposiums 1988*; herausgegeben von Paul Weingartner und Gerhard Schurz. Wien: Hölder-Pichler-Tempsky.
Metzinger, T. (1990). Kriterien für eine Theorie zur Lösung des Leib-Seele-Problems. *Erkenntnis*, **32**, 127–145.
Metzinger, T. (1991). Das Leib-Seele-Problem in den achtziger Jahren. *Conceptus*, **XXV**, Nr. 64, 99–114.

Metzinger, T. (1994). *Schimpansen, Spiegelbilder, Selbstmodelle und Subjekte. Drei Hypothesen über den Zusammenhang zwischen mentaler Repräsentation und phänomenalem Bewußtsein*. In: Krämer 1993.
Metzinger, T. (1995a) [Hrsg.]. *Bewußtsein – Beiträge aus der Gegenwartsphilosophie.* Paderborn: mentis.
Metzinger, T. (1995b). *Generelle Einleitung: Das Problem des Bewußtseins.* In Metzinger 1995a.
Metzinger, T. (1995c). *Ganzheit, Homogenität und Zeitkodierung.* In Metzinger 1995a.
Metzinger, T. (1995d). *Perspektivische Fakten? Die Naturalisierung des „Blick von nirgendwo".* In: Meggle, G./Nida-Rümelin, J. (1997) [Hrsg.], *ANALYOMEN 2 – Perspektiven der Analytischen Philosophie.* Berlin und New York: de Gruyter.
Metzinger, T. (1997). *Ich-Störungen als pathologische Formen mentaler Selbstmodellierung.* In: Northoff, G. [Hrsg.], *Neuropsychiatrie und Neurophilosophie.* Paderborn: mentis.
Metzinger. T. (1998). *Präsentationaler Gehalt.* In: Heckmann, H.-D./Esken, F. (Hrsg.), *Bewußtsein und Repräsentation.* Paderborn: mentis.
Metzinger, T. (1999). *Reflexives Selbstbewußtsein und die neurobiologische Verankerung der phänomenalen Erste-Person-Perspektive.* In: A. Newen/K. Vogeley (Hrsg.), *Das Selbst und seine neurobiologischen Grundlagen.* Paderborn: mentis.
Metzinger, T. (in Vorbereitung). *Der Traum – ein zweites Leben? Luzidität als Problem für eine philosophische Theorie des Geistes.*
Metzler, J./Shephard, R.N. (1974). *Transformational studies of the internal representation of three-dimensional objects.* In: Solso 1974.
Miller, S.D./Triggiano, P.J. (1992). The psychophysiological investigation of multiple personality disorder: review and update. *American Journal of Clinical Hypnosis,* **35,** 47–61.
Millikan, R.G. (1984). *Language, Thought, and other biological Categories.* Cambridge/MA: MIT-Press..
Milner, D./Rugg, M. (1992) [Eds]. *The Neuropsychology of Consciousness.* London Academic Press.
Minsky, M. (1975). *A Framework for representing Knowledge.* In: Winston 1975.
Mitchison, G. (1983). Siehe Crick/Mitchison 1983.
Munitz, M.K. (1971)[Ed]. *Identity and Individuation.* New York.
Myers, G.e./Irani, K.D. (1983)[Eds]. *Emotion: Philosophical Studies.* New York: Haven Publications.

Nadel, L. *et al.* (1989)[Eds]. *Neural Connections, Mental Computation.* Cambridge und London: MIT-Press (Bradford Books).
Nagel, T. (1965). Physicalism. *Philosophical Review*, **74**, 339–56; deutsch in: Bieri 1981.
Nagel, T. (1967). Review: Alan R. White, „Attention". *Philosophical Review*, 406–409.
Nagel, T. (1969). The Boundaries of Inner Space. *Journal of Philosophy*, 452–458.
Nagel, T. (1970). *The Possibility of Altruism.* Oxford University Press; nachgedruckt (1978): Princeton University Press.
Nagel, T. (1971). Brain bisection and the unity of consciousness. *Synthese*, **XX**, 396–413. Auch in Glover 1976; deutsch in: Nagel 1984.
Nagel, T. (1972). Review: D.C. Dennett, „Content and Consciousness". *Journal of Philosophy*, 220–224.
Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, **83**, 435–450; deutsch in: Bieri 1981 und Nagel 1984.
Nagel, T. (1979). *Mortal Questions.* Cambridge University Press. Deutsch (1992)[1984]: *Über das Leben, die Seele und den Tod.* Neue Wissenschaftliche Bibliothek. Athenäum Sonderausgabe. Frankfurt: Hain.
Nagel, T. (1980). *The Limits of Objectivity.* In: McMurrin 1980, 75–139. Deutsch (1991): Die Grenzen der Objektivität. Philosophische Vorlesungen. Übersetzt von Michael Gebauer. Stuttgart: Reclam.
Nagel, T. (1983). Review: Colin McGinn, „The Subjective View". *Times Literary Supplement*, **18**, November 1983.

Nagel, T. (1983). *The Objective Self*. Zuerst in: Ginet/Shoemaker 1983: 211–232. Eine erste deutsche Version findet sich in Siep 1983.
Nagel, T. (1986). *The View from Nowhere*. New York und Oxford: Oxford University Press. Deutsch (1992): *Der Blick von Nirgendwo*. Übersetzt von Michael Gebauer. Fankfurt: Suhrkamp.
Nagel, T. (1991). What we have in Mind when we say we're thinking (Rezension von D. Dennett: *Consciousness Explained). Wall Street Journal*, 7. 11. 1991.
Narayan, A. (1990). *On Being A Machine*. Volume 2: Philosophy of Artificial Intelligence. New York: Ellis Horwood.
Nemirow, L. (1979). *Functionalism and the subjective Quality of Experience*. Dissertation, Stanford University: University Microfilms International.
Nemirow, L. (1980). Review: Thomas Nagel, „Mortal Questions". *Philosophical Review*, **89**, 473–477.
Nemirow, L. (1990). *Physicalism and the cognitive Role of Acquaintance*. In: Lycan 1990.
Nyanaponika (1975). *Geistestraining durch Achtsamkeit*. Konstanz: Christiani.

Oakley, D.A. (1985)[Ed]. *Brain and Mind*. London: Methuen.
Oatley, K. (1985). *Representations of the physical and social world*. In: Oakley 1985.
Oatley, K. (1988). *On changing one's mind: a possible function of consciousness*. In: Marcel/ Bisiach 1988.
Oetke, C. (1988). *„Ich" und das Ich – Analytische Untersuchungen zur buddhistisch–brahmanischen Atmankontroverse*. Stuttgart.
Ornstein, R. (1986). *Multimind*. London.

Pagels, H. (1988). *The Dreams of Reason: The Computer and the Rise of the Sciences of Complexity*. New York: Simon & Schuster.
Palm, G./Aertsen, A. (1986) [Eds]. *Brain Theory*. Berlin: Springer.
Palmer, J. (1978). The Out-of-body-experience: A psychological Theory. *Parapsychology Review*, **9**, 19–22.
Papineau, D. (1987). *Reality and Representation*. Oxford: Blackwell.
Patzig, G., Scheibe, E. und Wieland, W. (1977)[Hrsg.]. *XI. Deutscher Kongreß für Philosophie, Logik–Ethik–Theorie der Geisteswissenschaften*. Hamburg: Meiner.
Peacocke, C. (1989). No Resting Place. *Philosophical Review,* 65–82.
Peirce, C.S. (1958)[1893]. *Collected Papers*. Cambridge/MA.
Pinker, S./Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed model of language acquisition. *Cognition,* **28**, 73–193.
Pitkin, H.F. (1967). *The concept of representation*. Berkeley. University of California Press.
Place, U.T. (1956). Is consciousness a brain process? *British Journal of Psychology*, **LXVII**; auch in: Borst 1970.
Place, U.T. (1988). Thirty years on – Is consciousness still a brain process? *Australasian Journal of Philosophy*, **66** (2), 208–219.
Podlech, A. (1984). *Repräsentation*. In: Brunner/Conze/Koseleck 1984.
Pollock, J.L. (1989). *How to build a Person: A Prolegomenon*. Cambridge und London: MIT-Press (Bradford Books).
Pöppel, E. (1985). *Grenzen des Bewußtseins*. München: DTV.
Pöppel, E. (1987). *Blindsight, Residual Vision*. In: Adelmann 1987, Band 1.
Pöppel, E. (1989)[Hrsg.].*Gehirn und Bewußtsein*. Weinheim: VCH Verlagsgesellschaft.
Pöppel, E. *et al.* (1973). Residual vision function after brain wounds involving the central visual pathways in man. *Nature*, **243**, 295f.
Popper, K. (1934b). *Logik der Forschung*. Wien: Julius Springer.
Popper, K. und Eccles, J. (1982). *Das Ich und sein Gehirn*. München: Piper.
Posner, M. I. (1989)[Ed]. *Foundations of Cognitive Science*. Cambridge/MA und London: MIT-Press.
Price, R./Cohen, D. (1988). *Lucid dream induction: An empirical evaluation*. In: Gackenbach/ Laberge 1988.

Puccetti, R. (1973). Brain bisection and personal identity. *British Journal for the Philosophy of Science*, **24**, 339–355.
Puccetti, R. (1981). The Case for Mental Duality / Author's Response. *Brain and Behavioral Sciences*, **4**, 93–99, 116–122.
Putnam, H. (1964). *Dreaming and depth grammar*. In: Butler 1962; nachgedruckt in Putnam 1975.
Putnam, H. (1975). *Mind, Language and Reality. Philosophical Papers, Volume 2*. Cambridge: Cambridge University Press.
Putnam, H. (1981). *Die Natur mentaler Zustände*. In: Bieri 1981.
Putnam, H. (1987). *Computational Psychology and Interpretation Theory*. In: Born 1987.
Putnam, H. (1991). *Repräsentation und Realität*. Frankfurt: Suhrkamp.
Pylyshyn, Z. (1973b). What the Mind's eye tells the Mind's Brain: A Critique of mental Imagery. *Psychological Bulletin*, **80**, 1–24.
Pylyshyn, Z. (1979b). The rate of Mental rotation of Images: A Test of a holistic Analogue Hypothesis. *Memory and Cognition*, **7**, 19–28.
Pylyshyn, Z. (1980). Cognition and Computation: Issues in the Foundations of Cognitive Science. *Behavioral and Brain Sciences*, **3** (1), 111–132.
Pylyshyn, Z. (1981). The Imagery Debate: Analogue Media versus tacit Knowledge. *Psychological Review*, **88**, 16–45.
Pylyshyn, Z. (1984). *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge und London: MIT-Press.

Ramsey, W., Rumelhart, D.E., Stich, S. (1990)[Eds]. *Philosophy and connectionist theory*. Hillsdale/NJ: Erlbaum.
Rehkämper, K. (1990). *Mentale Bilder – Analoge Repräsentationen*. In: Freksa/Habel 1990.
Rehkämper, K. (im Druck). Picture yourself in a boat on a river – Über die Bildhaftigkeit mentaler Repräsentationen. *Akten des XV. Deutschen* Kongresses für Philosophie in Hamburg 1990, herausgegeben von Herbert Schnädelbach.
Rheinwald, R. (1991b). Können Maschinen eine Sprache sprechen? Sind Computerprogramme syntaktisch oder semantisch? *Kognitionswissenschaft*, **2**, 37–49.
Ritter, H. (1988). *Selbstorganisierende neuronale Karten.* Dissertation. München (TU).
Ritter, H./Martinetz, T./Schulten, K. (1990). *Neuronale Netze – Eine Einführung in die Neuroinformatik selbstorganisierender Netzwerke*. Bonn, München, Reading: Addison-Wesley.
Rorty, A.O (1976)[Ed]. *The Identities of Persons*. Berkeley/Los Angeles: University of California Press.
Rorty, R. (1981a). *Leib–Seele Identität, Privatheit und Kategorien*. In: Bieri 1981.
Rorty, R. (1981b). *Unkorrigierbarkeit als Merkmal des Mentalen*. In: Bieri 1981.
Rorty, R. (1981c). *Der Spiegel der Natur*. Frankfurt: Suhrkamp.
Rosenberg, J. (1986). *The thinking Self.* Philadelphia.
Rosenthal, D. (1983). *Emotions and the Self*. In: Myers/Irani 1983.
Rosenthal, D. (1986). Two Concepts of Consciousness. *Philosophical Studies*, **49**, 3.
Rosenthal, D. (1990). *A Theory of Consciousness.* Report No. 40/1990 der Research Group on MIND AND BRAIN, Perspectives in Theoretical Psychology and the Philosophy of Mind, ZiF, Universität Bielefeld.
Rosenthal, D. (im Druck). *Thinking that one thinks*. In: Humphreys/Davies (im Druck).
Rumelhart, D. (1980). *Schemata: The building blocks of cognition*. In: Bruce/Brewer 1980.
Rumelhart, D.E./McClelland, J.L. & The PDP Research Group (1986)[Eds]. *Parallel distributed processing. Explorations in the microstructure of cognition*. Volume 1. Cambridge/MA: MIT Press.
Russell, B. (1940). *An Inquiry into Meaning and Truth*. London.
Ryle, G. (1949). *The Concept of Mind*. Hammondsworth: Penguin Books. Deutsch (1969): *Der Begriff des Geistes*. Stuttgart: Reclam.

Sacks, O. (1987). *Der Mann, der seine Frau mit einem Hut verwechselte*. Reinbek: Rowohlt.
Savage, C.W.(1978)[Ed]. Perception and Cognition. Issues in the Foundations of Psychology. *Minnesota Studies in the Philosophy of Science. Vol. 9*. Minneapolis.
Schank, R./Abelson, R. (1977). *Scripts, Plans, Goals and Understanding*. Hillsdale/NJ.
Scheerer, E. (1990a). *Mental Representation: Its History and present Status – 1. ‚Representatio' from Cicero to Suarez*, Report No. 27/1990 der Research Group on MIND AND BRAIN, Perspectives in Theoretical Psychology and the Philosophy of Mind, ZiF, Universität Bielefeld.
Scheerer, E. (1990b). *Mental Representation: Its History and present Status – 2. Descartes, his Followers and his Opponents*, Report No. 43/1990 der Research Group on MIND AND BRAIN, Perspectives in Theoretical Psychology and the Philosophy of Mind, ZiF, Universität Bielefeld. Vgl. auch: Artikel *Repräsentation I* und *IV* im Historischen Wörterbuch der Philosophie.
Schmitt, F.O./Worden, F.G. (1974). *The Neurosciences Third Study Program*. Cambridge/MA und London: MIT-Press.
Schnelting, K. (1984). *Hilfe ich träume*. München.
Schnotz, W. (1988). *Textverstehen als Aufbau mentaler Modelle*. In: Mandl/Spada 1988.
Scholz, O. (1991a). *Bild, Darstellung, Zeichen – Philosophische Theorien bildhafter Darstellung*. Freiburg i. Br. und München: Alber.
Scholz, O. (1991b). Artikel *Repräsentation III (19. und 20. Jahrhundert)* im Historischen Wörterbuch der Philosophie.
Schütt, H.–P. (1981). *Einleitung* zu: „Der Begriff der Person". In: Bieri 1981, Vierter Teil.
Searle, J. (1980). Minds, Brains, and Programs. *Behavioral and Brain Sciences*, **3**, 417–424.
Searle, J. (1986). *Geist, Hirn und Wissenschaft. Die Reith Lectures 1984*. Frankfurt: Suhrkamp.
Searle, J. (1992). *The Rediscovery of the Mind*. Cambridge/MA, London: MIT-Press.
Sellars, W. (1963). *Science, Perception and Reality*. London: Routledge and Kegan Paul.
Shallice, T. (1972). Dual functions of consciousness. *Psychological Review*, **79**, 383–393.
Shallice, T. (1988). *Information-processing Models of Consciousness: Possibilities and Problems*. In: Marcel/Bisiach 1988.
Shephard, R.N./Chipman, S. (1970). Second-order Isomorphism of internal Representations: Shapes of States. *Cognitive Psychology*, **1**, 1–17.
Shephard, R.N./Cooper, L.A. (1982). *Mental images and their transformations*. Cambridge/MA: MIT Press.
Shephard, R.N. und Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, **171**, 701ff.
Sher, G. (1977). Kripke, Cartesian intuitions, and materialism. *Canadian Journal of Philosophy*, **7**, 227–238.
Shoemaker, S. (1975). Functionalism and Qualia. *Philosophical Studies*, **27**, 291–315.
Shoemaker, S. (1981a). Functionalism, qualia, and intentionality. *Philosophical Topics*, **12**, 93–118.
Shoemaker, S. (1981b). *Selbstreferenz und Selbstbewußtsein*. In: Bieri 1981.
Shoemaker, S. (1981c). Absent qualia are impossible, a reply to Block. *Philosophical Review*, **XC**, 581–599.
Shoemaker, S. (1982). The Inverted Spectrum. *Journal of Philosophy*, **79**, 357–381.
Shoemaker, S. (1983). Siehe Ginet/Shoemaker 1983.
Siegel, R.K./Jarvik, M.E. (1975). *Drug-induced hallucinations in animals and man*. In: Siegel/West 1975.
Siegel, R.K./West, L.J. (1975)[Eds]. *Hallucinations. Behavior, experience, and theory*. New York: Wiley.
Siep, L. (1983)[Ed]. *Identität der Person – Aufsätze aus der nordamerikanischen Gegenwartsphilosophie*. Basel/Stuttgart: Schwabe & Co. AG.
Sinclair, R./Jarvella, R.J./Levelt, W.J.M. (1978)[Eds]. *The child's conception of language*. Berlin: Springer.
Singer, W. (1989). *Zur Selbstorganisation kognitiver Strukturen*. In: Pöppel 1989.
Singer, W. (1990). Search for coherence: a basic principle of cortical self-organization. *Concepts in Neuroscience*, **1**, 1–26.

Singer, W. (1993). Synchronization of cortical activity and its putative role in information processing and learning. *Annual Review of Physiology*, **55**, 349–374.
Singer, W. (1995). *The organization of sensory motor representations in the neocortex; A hypothesis based on temporal coding.* In: Umilta/Moscovitch (im Druck).
Smart, J. (1959). Sensations and brain processes. *Philosophical Review*, **LXVIII,** auch in: Borst 1970.
Smart, J. (1961). Further remarks on sensations and brain processes. *Philosophical Review*, **LXX**. Auch in: Borst 1970.
Smart, J. (1962). Brain processes and incorrigibility. *Australasian Journal of Philosophy*, **LX**. Auch in: Borst 1970.
Smart, J. (1963a). *Philosophy and scientific realism.* London.
Smart, J. (1963b). Materialism. *Journal of Philosophy*, **LX,** auch in: Borst 1970.
Smolensky, P. (1987). Conectionist AI, and the brain. *Artificial Intelligence Review*, **1**, 95–109.
Smolensky, P. (1988). On the proper treatment of connectionsism. *Behavioral and Brain Sciences*, **11**, 1–74.
Sober, E. (1985). Panglossian Functionalism and the Philosophy of Mind. *Synthese*, **64**.
Solso, R.L. (1974)[Ed]. *Theories in cognitive psychology: The Loyola Symposium*. Hillsdale/NJ: Erlbaum.
Sparrow, G. (1976). A personal testimony: Developing lucidity in my dreams. *Sundance Community Dream Journal*, **1**, 4–17.
Sperry, R.W. (1970). Perception in the Absence of Neocortical Commissures. In: *Perception and its Disorders*, Res.Publ. A.R.N.M.D., **48**, 123–138.
Sperry, R.W. (1974). *Lateral Specialization in the surgically separated Hemispheres.* In: Schmitt/Worden 1974.
Stanton, W.L. (1978). *Anomalous Monism and the Mental qua Mental.* Dissertation. Princeton.
Stegmüller, W. (1978). *Hauptströmungen der Gegenwartsphilosophie.* Band 1. Stuttgart: Kröner.
Steiner, G. (1988). *Analoge Repräsentationen*. In: Mandl/Spada 1988.
Stephan, A. (1999). *Emergenz: Von der Unvorhersagbarkeit zur Selbstorganisation.* Dresden/München: Dresden University Press.
Stephan, A. (im Druck). Jaegwon Kim – ein Lotse zwischen der Skylla des Dualismus und der Charybdis des Reduktionismus. *Akten des XV. Deutschen Kongresses für Philosophie* in Hamburg 1990, herausgegeben von Herbert Schnädelbach.
Sterelny, K. (1986). The Imagery Debate. *Philosophy of Science*, **53**(4), 560–583. Auch in: Lycan 1990.
Stewart, K. (1972). *Dream theory in Malaya.* In: Tart 1975.
Stich, S. (1978). Autonomous Psychology and the Belief-Desire Thesis. *The Monist*, **61**.
Stich, S. (1983). *From Folk Psychology to Cognitive Science*. Cambridge/MA: MIT-Press.
Stich, S. (1992). What is a Theory of Mental Representation? *Mind*, Vol. 101, **402**, 241–261.
Stoecker, R. (1992). *Was sind Ereignisse? Eine Studie zur analytischen Ontologie.* Berlin/New York: Walter de Gruyter.
Stoerig, P. *et al.* (1985). Signal detection analysis of residual vision in a field defect due to a post-geniculate lesion. *Neuropsychologia*, **23**, 589–599.
Stoerig, P./Cowey, A. (1990). Wavelength Sensitivity in Blindsight. *Nature*, **342**, 916–918.
Stoerig, P./Cowey, A. (1991a). Increment-threshold spectral sensitivity in blindsight. Evidence for colour opponency. *Brain*, **114**, 1487–1512.
Stoerig, P./Cowey, A. (1992). Wavelength discrimination in blindsight. *Brain*, **115**, 425–444.
Stoerig, P./Cowey, A. (1993). *Blindsight in Perceptual Consciousness: Neuropsychological Aspects of Striate Cortical Function.* In: Gulyas/Oltoson/Roland 1993.
Stoyan, H. (1986)[Hrsg.]. *GWAI 85. 9. German Workshop on Artficial Intelligence*, Informatik Fachberichte **118**, Berlin: Springer
Strawson, P.F. (1972). *Einzelding und logisches Subjekt (Individuals).* Stuttgart: Reclam.

Tart, C. (1967). The control of nocturnal dreaming by means of posthypnotic suggestion. *International Journal of Parapsychology*, **9**, 184–189.
Tart, C. (1972). *The „High" Dream: A New State of Consciousness.* In: Tart 1972.
Tart, C. (1972)[Ed]. *Altered States of consciousness: A book of readings.* New York: Anchor, Doubleday. (1969: New York, Toronto, London – John Wiley).
Tart, C. (1975). States of consciousness and state-specific sciences. *Journal of Altered States of Consciousness*, **2**(1).
Tart, C. (1988). *From spontaneous event to lucidity – A review of attempts to consciously control nocturnal dreaming.* In: Gackenbach/Laberge 1988.
Taylor, C. (1987). *Overcoming Epistemology.* In: Baynes/Bohman/McCarthy 1987.
Taylor, J. G./Caianello, E. R./Cotterill, R. M. J./Clark, J. W. (1992) [Eds]. *Neural Network Dynamics.* London: Springer.
Teuber, H. (1965). Postscript: Some needed revisions of the classical views of agnosias. *Neuropsychologia*, **3**, 371–378.
Tholey, P. (1982). Bewußtseinsänderung im Schlaf: Wach' ich oder träum' ich? *Psychologie Heute*, **9**(12).
Tholey, P. (1983). Techniques for inducing and manipulating dreams. *Perceptual and Motor Skills*, **59**, 875–878.
Tholey, P. (1984). *Der Klartraum – Hohe Schule des Traums.* In: Schnelting 1984.
Tholey, P. (1987). *Schöpferisch träumen.* Niedernhausen: Falken Verlag.
Tholey, P. (1988). *A model for lucidity training as a means of self-healing and psychological growth.* In: Gackenbach/Laberge 1988.
Tranel, D./Damasio, A. (1985). Knowledge without awareness: An autonomic index of recognition by prosopagnosics. *Science*, **228** (21), 1453f.
Tugendhat, E. (1979). *Selbstbewußtsein und Selbstbestimmung – Sprachanalytische Interpretationen.* Frankfurt: Suhrkamp. (4. Auflage 1989)
Turing, A.M. (1937). On computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society, 42*, 230–265.
Turing, A.M. (1950). Computing Machinery and Intelligence. *Mind*, **59**. Deutsch: *(Kann eine Maschine denken?)* in: Enzensberger 1967.
Tye, M. (1986). The Subjectivity of Experience. *Mind*, **95**.

Umilta, C./Moscovitch, M. (Eds). *Attention and Performance XV: Conscious and Nonconscious Information Processing.* Cambridge/MA: MIT-Press.

Van Gulick, R. (1985). Physicalism and the Subjectivity of the Mental. *Philosophical Topics*, Vol. XIII, **3**, 51–70.
Van Gulick, R. (1988a). A Functionalist Plea for Self-Consciousness, *The Philosophical Review*,Vol. XCVII, **2**, 149–181.
Van Gulick, R. (1988b). *Consciousness, intrinsic Intentionality, and self-understanding Machines.* In: Marcel/Bisiach 1988.
Villanueva, E. (1990)[Ed]. *Information, Semantics and Epistemology.* Oxford: Blackwells.
Vollmer, G. (1992). Die Vierte bis Siebte Kränkung des Menschen – Gehirn, Evolution und Menschenbild. *Gießener Diskurse*, **6/7**.
von der Malsburg, C. (1981). *The correlation theory of brain functioning.* Internal Report 81–2. Göttingen: Max-Planck-Institut für Biophysikalische Chemie.
von der Malsburg, C. (1986). *Am I thinking assemblies?* In: Palm/Aertsen 1986.
von Eckardt, B. (1984). Cognitive Psychology and principled Skepticism. *Journal of Philosophy*, **81**, 67–88.

Weiskrantz *et al.* (1974). Visual capacity in the hemianopic field following a restricted occipital ablation. *Brain*, **97**, 709ff.
Weiskrantz, L. (1968)[Ed]. *Analysis of behavioral change.* New York: Harper and Row.
Weiskrantz, L. (1972). Behavioral analysis of the monkey's visual nervous system. *Proceedings of the Royal Society*, **B 182**, 427–455.

Weiskrantz, L. (1986). *Blindsight: A case study and implications.* Oxford: Oxford University Press.
Weiskrantz, L. (1988). *Some contributions of neuropsychology of vision and memory to the problem of consciousness.* In: Marcel/Bisiach 1988.
Weissbord, D. (1989)[Ed]. *Mind, Value, and Culture: Essays in Honor of E.M. Adams.* Ridgeview Publishing.
Werth, R. (1983). *Bewußtsein – Psychologische, neurobiologische und wissenschaftstheoretische Aspekte.* Berlin: Springer.
Whitman, R., Ornstein, P. & Baldridge, B. (1964). An experimental approach to the psychoanalytic theory of dreams and conflicts. *Comprehensive Psychiatry*, **5**, 349–363.
Wilkes, K.V. (1984). Is Consciousness important? *British Journal for the Philosophy of Science*, **35**, 224–243.
Wilkes, K.V. (1988a). *Real People – Personal Identity without Thought Experiments.* Oxford: Clarendon Press.
Wilkes, K.V. (1988b). ––––, *yishi, duh, um, and Consciousness.* In: Marcel/Bisiach 1988.
Winson, J. (1991). Neurobiologie des Träumens. *Spektrum der Wissenschaft*, **1**/91, 126–134.
Winston, P. (1975)[Ed]. *The Psychology of Computer Vision.* New York.
Wittgenstein, L. (1970)[1918]. *Tractatus logico-philosophicus.* Werkausgabe Band 1. Frankfurt: Suhrkamp.
Wittgenstein, L. (1970)[1958]. *Das Blaue Buch.* Werkausgabe Band 5. Frankfurt: Suhrkamp.
Wolman, B. (1979)[Ed]. *Handbook of Dreams.* New York.

Zihl, T. (1980). ‚Blindsight': Improvement of visually guided eye movements by systematic practice in patients with cerebral blindness. *Neuropsychologia*, **18**, 71–77.

6.2 Personenverzeichnis

Abelson, R. 123
Abrahamsen, A. 63, 70, 125f, 180
Aghajanian, G.K. 145
Albert, H. 250
Allport, A. 42, 88
Alvarado, C. 187
Anaximenes 187
Anderson, J.A 71
Anscombe, G.E.M 232
Anton, G. 143
Aristoteles 187f, 198
Armstrong, D.M. 24, 211, 271
Arnauld 98, 109

Baars, B. 98
Barinaga, M. 92, 106
Bartels, A. 11, 24, 237
Bechtel, W. 63, 70, 125f, 180
Beckermann, A. 28, 31, 53, 80, 89, 216
Benson, D.F. 143
Bergson, H. 75
Bieri, P. 11, 24, 29, 42, 63, 76, 80, 88, 90, 118, 130, 143, 172, 187, 211, 213ff, 216ff, 218, 239, 269, 290
Bisiach, E. 72, 90f, 93, 104, 139, 141
Blackmore, S. 187f, 197, 200, 203
Blau, U. 23, 57, 236
Block, N. 50, 53, 58, 93, 214, 238
Boden, M. 28, 58
Bodis-Wollner, I. 141
Boër, S. 233
Bogen, J. 31
Born, R. 275
Borst, C.V. 24
Braitenberg, V. 159
Brentano, F. 17f, 127
Brewer, W. 124
Bunge, M. 89, 95
Burman 50
Buschlinger, W. 226
Butterfield, J.F. 78, 219

Campbell, K. 214
Casey, J. F. 191
Chipman, S. 50, 110
Churchland, P.M. 33, 36f, 45, 50, 71, 79, 96, 130f, 150, 158, 180, 211f, 214, 218, 237f, 261, 286
Churchland, P.S. 60, 67, 72, 91, 130, 140, 211, 261
Cicero 49
Clark, A. 63, 80, 98, 125, 170, 173ff
Cooper, L.A. 111
Cowey, A. 141f, 259
Craik, K.J.W. 50, 118, 121f, 189f
Crane, H. 85
Crick, F. 92, 107, 147, 149, 201
Cummins, R. 69, 96, 100, 208
Curren, R. 196
Cutler, G. 196
Cytowic, R.E. 86

Damasio, A. 138, 143
Danto, A. 85
Davidson, D. 25, 28, 32, 48, 96, 118, 143
Davis, L. 238
Dennett, D. 29, 32, 48, 50, 53, 74ff, 79, 81, 85, 93ff, 103, 128, 130, 138, 140, 142, 172, 190ff, 200f, 205, 211, 214ff, 216, 224, 237f
Dentan, R.K. 197, 201
Descartes, R. 25, 30f, 50, 74, 88, 154, 212, 218, 230, 247, 261
Dijk, W. 124
Diogenes von Apollonia 187
Dittrich, A. 182ff, 217
Dretske, F. 67, 69, 131f
Dreyfus, H.L. 275
Dreyfus, S.E. 275

Eccles, J.C. 13, 31, 72, 89, 230, 277
Edelman, G.E. 206, 286
Eimer, M. 53, 96, 100, 122, 208
Engel, A. 92, 106

Farrell, B. 212
Feigl, H. 24
Feyerabend, P. 214
Fichte, I. G. 99, 223
Finke, R.A. 50, 117
Flohr, H. 98, 91f, 104, 156
Fodor, J. 53, 63, 67, 79, 109, 119, 125f, 174, 201, 214, 258, 268
Foss, J. 211, 220, 233
Frank, M. 27, 100, 223, 225, 229
Frankfurt, H.G. 274
Freud, S. 54, 156, 184

Gackenbach, J. 196ff
Garfield, P. 196
Gassendi 50
Gazzaniga, M.S. 31
Gebauer, M. 211, 228
Gentner, D. 124
Ginet, C. 234
Globus, G.G. 90

Goschke, T. 63, 70, 111, 125f, 175, 258
Gray, C. 92, 107
Green, C. 198
Greenberg, J.P. 143
Gregory, R. 156
Griffiths, P. 256

Hardin, C.L. 76, 78, 83f, 113, 122, 142, 219
Hastedt, H. 89
Haugeland, J. 32, 95, 202, 243
Hebb, D. 91
Hegel, G. W. F. 14, 188
Heidegger, M. 75
Heimann, H. 181
Helm, G. 48, 63, 70f, 111f, 117, 119, 126
Henrich, D. 99
Herken, R. 28
Herrmann, T. 51
Hildegard von Bingen 184
Hill, C.S. 237
Hinton, G. 70f, 126, 201
Hobson, A. 146ff, 201, 203
Hofstadter, D. 211
Hopfield, J.J. 181
Horgan, T. 48
Horstmann, R.-P. 208
Howland, F. 191
Humphrey, N. 190ff
Hurvich, L.M. 113
Husserl, E. 127, 228, 280

Irwin, H. 187f, 197

Jackendoff, R. 36, 55, 59, 65, 79, 94, 107f
Jackson, F. 37, 76, 78f, 81f, 211, 214, 218ff, 238f
Jacobs, B.L. 145
Jarvik, M. E. 183
Johnson-Laird, P.N. 58, 61, 63, 70, 93, 105, 112, 122ff, 126, 130, 135

Keane, M.T. 58
Kennedy, M.B. 91
Kiefer, E. 159, 276
Kim, J. 32, 48, 89, 95, 172, 202
Kinsbourne, M. 30, 170
Kintsch, A. 124
Kirk, R. 214
Kitcher, P. 238, 261
Koch, C. 92, 107
Koppelberg, D. 63, 70, 111, 125f, 175, 258
Kosslyn, S.M. 50, 57f
Kreiter, T. 92, 107
Kripke, S.A. 24, 39, 167, 233, 237, 263, 271
Kukla, R. 69
Kurthen, M. 73, 98, 217

LaBerge, S. 196ff
Langton, C.G. 92
Lanz, P. 32
LeDoux, J.E. 31
Leibniz 50
Leplin, J. 237
Leuner, H.-C. 145
Levelt, W.J.M.
Levin, J. 45, 82, 237
Lewis, D. 24, 76, 131, 211, 219, 238
Libet, B. 72, 102, 153, 286
Locke, D. 214
Lycan, W.G. 24, 35, 411, 50, 53, 55, 59, 88, 93, 102, 107, 211, 214, 222, 233f, 236ff, 261

Mackay, D.M. 95
Malcolm, N. 206f, 211, 230, 232
Malebranche 98f
Maloney, J.C. 32, 211, 214
Mandler, G. 95
Marcel, A. 91, 107
Marr, D. 110, 122
Marshall, J.C. 95
Martinetz, T. 110
Maxwell, G. 90
McCarley R.W. 147, 201
McClelland, J.L. 70f, 93, 126
McGinn, C. 24, 67, 105, 115, 121f, 130, 133, 164, 211, 237, 268
McMullen, C. 211, 237
Meinong, A. 75
Meister Eckehart 184
Melzack, R. 160, 187, 261
Mesulam, M-M. 139
Metzinger, T. 24f, 31f, 80, 89, 118, 166, 169, 198, 202, 216, 222, 237, 269, 279
Metzler, J. 111
Millikan, R.G. 67
Minsky, M. 123
Mitchison, G. 147, 149, 201
Morton, J. 95

Nadel, L. 71
Nagel, T. 10, 23, 31, 37, 39, 44f, 58, 82, 128, 152, 155, 166ff, 179, 183, 185, 187, 200, 210-240, 248, 252, 254ff, 264, 273, 277, 282, 288, 289
Narayan, A. 28
Nemirow, L. 75, 79, 211, 219
Nyanaponika 101

Oatley, K. 163, 171, 175, 189f, 208
Ockham, W. 99
Oetke, C. 256
Ornstein, R. 158

Palmer, J. 188, 197
Papineau, D. 67
Peacocke, C. 211
Peirce, C.S. 250
Piantanida, T.P. 85
Place, U.T. 24, 41, 271
Podlech, A. 49
Pollock, J.L. 172
Pöppel, E. 74, 77, 87, 102, 141f, 181, 259
Popper, K. 19, 21, 25, 29, 72, 89f, 230, 250, 277
Puccetti, R. 31
Putnam, H. 28, 53, 121, 198, 200, 276
Pylyshyn, Z. 58, 63, 108, 125f, 174, 258

Rehkämper, K. 50, 58, 110f, 115, 117
Ritter, H. 110
Rolland, R. 184
Rorty, A.O 143
Rorty, R. 85, 212, 214, 269
Rosenberg, J. 249
Rosenfeld, E. 71
Rosenthal, D. 55, 65, 94, 98, 162
Rumelhart, D.E. 70f, 93, 123, 126, 175
Russell, B. 130
Ryle, G. 64, 214, 216

Sacks, O. 84f, 139ff, 177ff
Schank, R. 123
Scheerer, E. 49f, 98f, 110
Schlick, M. 216
Schnotz, W. 124, 130f
Scholz, O. 49f, 121f, 169
Schulten, K. 110
Schütt, H.-P. 172
Searle, J. 28, 206
Sejnowski, T.J. 60, 201
Sellars, W. 73, 102, 217
Seuse 184
Shallice, T. 93, 95
Shephard, R.N. 50, 110f
Sher, G. 237
Shoemaker, S. 214, 234, 238, 253
Shwartz, S.P. 58
Siegel, R.K. 116, 145, 183
Siep, L. 172, 211
Singer, W. 92, 104, 107
Smart, J. 24, 35, 271
Smolensky, P. 63, 70f, 126, 174f
Sober, E. 93
Sparrow, G. 196
Spencer, H. 75
Sperry, R.W. 31
Spinoza 228
Stanton, W.L. 25
Stegmüller, W. 24, 127
Steiner, G. 64, 111, 124
Stephan, A. 48
Sterelny, K. 58
Stevens, A.L. 124
Stewart, K. 197, 201
Stich, S. 49, 67, 130, 261
Stoecker, R. 202
Stoerig, P. 141f, 259
Strawson, P.F. 132, 172, 216, 266

Tart, C. 194
Tauler 184
Taylor, C. 237
Teuber, H. 138
Tholey, P. 195, 197
Tranel, D. 138
Trulson, M.E. 145
Tugendhat, E. 99f
Turing, A.M. 28, 121
Tye, M. 211

Vallar, G. 139
Van Gulick, R. 63, 99, 115, 155, 173, 211, 245, 251, 262, 281
Vollmer, G. 156, 285
von der Malsburg, C. 91, 107
von Eckardt, B. 261

Weiskrantz, L. 91, 107, 141f, 259
Weissbord, D. 59
Werth, R. 90, 141
Wilkes, K.V. 27, 31, 42, 87f, 142, 191, 226, 231, 238, 269
Wilson, D.H. 31
Winson, J. 149
Wittgenstein, L. 216, 228, 253

Zihl, T. 141

6.3 Begriffsverzeichnis

Abbildungsrelation 110
Absent-Qualia-Argument 214
Abstrakta
– und Prototypen 61
– abstrakte Information 68
– qualitative Erfassung 84, 270
– rationalisierende Verhaltenserklärungen 290
– Beschreibungsebenen 93, 270
– abstrakte Instrumente 129ff, 159ff, 173, 201 (siehe auch *Organe*)
– abstrakte Automaten 165, 276, 287
– Abstraktion/Introspektion 233
– als Bewußtseinsinhalte 263f, 270
– *als* Konkreta 270, 288
– abstrakte Räume 288
Agenten
– kognitive 28
– moralische 153
– mentale 166
– epistemische 268f
Agnosien 56, 137ff, 177ff
Ähnlichkeit
– allgemein 82, 121f, 129ff, 268, 290
– Selbstähnlichkeit 153ff, 159
Ähnlichkeitskriterien
– für interne Netzwerkrepräsentate 52, 71, 126
– repräsentationale und physikalische 68
Aktivierungsvektoren 52, 69, 107, 117, 288
Aktivierungszustände 51, 260, 289
Alltagsidiom 18, 21, 97, 238
Alltagsphänomenologie 95, 238, 246
Alltagspsychologie
– allgemein 17, 101, 138, 204, 218, 260ff
– *common-sense*-Taxonomie *mentaler* Zustände 27, 260ff
– „gelehrte“ 50
– Vorannahmen 87
Amnesien 180, 231, 278
Analog-Indikatoren 61, 258ff, 263ff, 278f
Analogien
– transmodale 78, 83
Analyse
– funktionale 28, 75ff
Analytizität 167, 248
Anomalie des Mentalen 25
Anomalien der psychischen Realität 137ff
Anomien 138
Anosognosien 104, 143
Anthropologie 14
– philosophische 16, 284ff, 291
– philosophisch-psychologische 54, 291
– naturalistische 230
Antons Syndrom 143f
Artefakte
– emotionale 81
– iatrogene 193
– innere Objekte 236
– psychische 259
Assoziationsfluß 181
Asymetrie
– epistemische 225
Auflösungsvermögen
– zeitliches 65
Aufmerksamkeit
– innere 55f, 101, 262
– Fokus 90
– neurobiologische Korrelate 106
– als Modell der Repräsentationsbeziehung 128f
– Störungen 139ff
Aufwachen
– allgemein 34, 200
– „falsches Erwachen“ 202
außerkörperliche Erfahrungen 187ff

Bedeutung
– innerer Zustände 17
– von Termen einer wissenschaftlichen Psychologie 39
– von Diskursen 123f
– Bedeutungsfunktion 233ff
Begriffe
– und Metaphern 126
– Bildung 131, 180
Begriffsgeschichte 13
Beschreibungsebenen
– allgemein 30, 271
– subpersonale 32, 48, 120, 168
– personale 48, 179, 205, 216, 285
– und Zuschreibungsbedingungen für „bewußt“ 90, 95
– zentriertes 9, 23ff, 222ff, 237, 265
– inkommensurable 96
– der mentalen Modellierung 120
– repräsentationale 208
– und Selbstreferenz 232
– abstrakte 246
Betriebssystem 93
bewußt
– als zweistelliges Prädikat 17f
– als einstelliges Prädikat 27f, 36
– Elemente des Begriffs 21ff, 50
– phänomenologische Hauptaspekte 22ff

Bewußtsein
– zentriertes 9, 23ff, 222ff, 237, 265
– von etwas 17
– allgemein 20f, 31, 42, 69, 72, 87ff, 259
– Gegebenheit des 25, 268
– Homogenität des 25, 30f, 88ff, 101,
– Konkretheit des 25, 288
– Inhalte 27, 47ff, 87, 115, 136ff
– von Maschinen 28f, 162
– künstliches 28f, 207
– Realität des 29
– Feldqualität des Erlebnisraums 30
– als distribuierte Netzwerkeigenschaft 30, 258
– Unteilbarkeit 30f
– Atomizität 31
– Selbsttransparenz des 31, 143, 218, 232, 239, 199, 269f
– reines 34
– subjektives 48, 65, 166ff, 272ff
– phänomenales 57, 122f, 258
– zeitliche Unmittelbarkeit 76
– als Explanandum/fehlender Begriff 88, 91ff, 104, 115, 207, 213ff
– nicht-fragmentiertes 88ff, 281
– als Hintergrundvariable 90
– repräsentationalistischer Begriff 97ff
– Modularität des computationalen 104
– neurobiologische Korrelate 106
– „Bewußtseinhelligkeit“ 139
– inneres sensorisches 157ff
– nicht-subjektives 184f
– aperspektivisches 184, 237, 256, 264ff, 281
– multizentriertes 193, 278
– der eigenen Entscheidungsfreiheit 195
– kritisches 195ff
– phänomenales 207
– *„Wie es ist, ...“* 214ff
– dezentriertes 237
– ohne Qualia 263ff
– holistisches 281, 287
Bilder
– hypnagoge 113, 182
Biographie
– innere 9, 175, 180, 192, 203
– psychische 16, 180
– psychische Historizität 275
biologische Imperative 162, 270, 285, 289
Blick von Nirgendwo
– allgemein 225ff, 254ff, 280ff, 288
– *Großer Blick von Nirgendwo* 185, 225f, 254, 281
blinder Fleck 85
Blindsicht 87, 137ff
borderline-Psychose 191
bottom-up-constraints 60, 206, 275
Brennpunkt
– des Erlebnisraums 23

Cambridge-Theorie der mentalen Repräsentation 121ff, 268
cartesianische Bühne 129
cartesianische Katastrophe 27
cogitationes 154, 235, 258
Cogito 154f, 161, 164, 275
Computational Mind 65, 94
Computational Theory of Mind 55
Computer-Metapher 19, 121, 285
Cyberspace 243

Datenstrukturen
– mentale 58
– in Modulen 79
– abstrakte Eigenschaften 83ff
– analoge 109
– unterdeterminierte 116, 186
– intendierte Klasse 208
– Einbettung 266ff
Deixis 279
Denken 70, 125ff, 164f, 180, 250, 272
Denkstörungen 181f
Depersonalisation 185
Derealisation 136
Designatoren
– starre 24, 271f
Diskonnektions-Syndrome 108
Dissoziationen
– funktionale 107
Doppelgänger-Intuition 214
Drei-Welten-Lehre 230
Dritte-Person-Perspektive 96
Dualismus
– ontologischer 18f
– interaktionistischer 19, 172

Ego
– epistemische Beziehungen zur Welt 50
– Egologie 99
– als Gedachtes 154f
– transzendentales 226, 228, 255f, 282
– als Individuum 271
– Transzendierung 281
Eidetiker 75, 113
Eigennamen 24, 270ff
Eigenreize 114, 160, 177ff
Eigenschaften
– psychologische 21, 53, 60, 100, 103, 109, 117, 121, 131, 172ff, 194, 204ff, 210f, 225, 235, 246, 248, 250
– Information über physische 81

– abstrakte, von Datenstrukturen 83ff, 102ff, 270
– abstrakte, von informationsverarbeitenden Systemen 92, 166
– repräsentationale und abstrakte 103, 258f, 270, 278
– relationale 110, 205ff, 230
– Verknüpfung physischer und intentionaler 129ff
– kausale 160
– epistemologische 165
– physikalische 172
– phänomenale 174, 259
– private 213ff, 232, 248, 260ff, 269f
– öffentliche 223ff, 231f, 248
– Kontingenz 217
– erster Ordnung 236
– nicht-relationale 260ff
– Mengen als Objekte 284
Eigenschaftsbindung
– repräsentationale 91, 98, 283
– durch phasengleiche neuronale Oszillationen 91, 106
– „von oben“ 107
Eigenwahrnehmung 157ff, 177ff, 187, 242
Einbettung
– mentaler Repräsentate ineinander 61f, 135ff
– mentaler Modelle ins Selbstmodell 205ff, 268ff
Einstellungen
– propositionale 123, 130f, 198, 250, 265, 283
Emergenzbegriff
– allgemein 89, 95
– symmetrische Typ-Emergenz 89
Emotionen 158, 162f, 184f, 193, 274f
empirische Forschung
– Rolle der 15
Emulation
– phänomenale 214, 220, 278f
Endlosschleifen 63
Energieerhaltungssatz 25
Entitäten
– nicht-physikalische 89, 107
– extramentale 99
– theoretische 117, 228
– logisch primitive 172
– metaphysische 213
– abstrakte 270
– ontologisch neutrale 276
Entkoppelungen
– phänomenale 107
Episodizität
– subjektiven Bewußtseins 34, 256
Erkenntnisfortschritt 26, 291
Erkenntnisideal
– wissenschaftliches 26
Erkenntnistheorie
– naturalistische 37f, 268
– natürlicher Repräsentationssysteme 173
Erklärungen
– kausale 15 (siehe auch *kausale Verhaltenserklärungen*)
– Instantiierungserklärungen 15 (siehe auch *Instantiierungserklärungen*)
Erlebnisgehalt
– allgemein 20, 26
– Asymmetrie zu theoretischem Gehalt 176
Erlebniskern 78
Erlebnisqualität
– der Inhalte des Selbstbewußtseins 21
Erlebnisraum 77, 287
Erlebnisse
– allgemein 35
– subjektive 23, 55
– von Personen 31, 34, 48
– aperspektivische 34, 184f, 281
– gehaltfreie 34
– subjektive Taxonomie 36
– zeitliche Unmittelbarkeit 74
Erlebnissubjekt
– allgemein 32, 68, 103, 122, 128f, 136, 210f, 244, 284
– Verkörperung 157ff, 264
– Begriff 207, 216, 266
Erlebnisträgerschaft 21
Erlebniswelten
– virtuelle 65
Essenz
– phänomenale 78
être répresentatif 99
Evidenz 29, 77, 269f
evolutionärer Kontext 59, 288

Fakten
– interne 52, 251ff
– perspektivische 221ff, 251ff, 282ff
– Erste-Person-Fakten 223ff, 236, 251ff
Fantasien 65
Farbensehen 81
Farbprädikate 86
Fehlschluß
– phänomenologischer 100, 161
– genetischer 247
– naiv-realistischer 281
Feld
– phänomenales 76, 89
– Feldqualität des Bewußtseins 90, 102
Folgerelationen 129, 165
folk psychology 50, 94, 260f

Formate
– von mentalen Repräsentaten 57, 131, 165ff, 169ff
– Integration 64, 107ff, 124ff
– von Präsentaten 79ff, 258ff, 274, 279f
– interne 80
– als Mengen von abstrakten Eigenschaften 80, 278
– Inkompatibilität 80
– Bewußtsein 102ff
– propositionale 125ff
– analoge und Einbettbarkeit 133
– von Selbstmodellen 159
– Formkonstanten von Halluzinaten 183
Frames 123f
Funktionalismus 19
– Maschinenfunktionalismus 53, 276
– Zwei-Ebenen-Funktionalismus 53
– Mikrofunktionalismus 80, 97, 121, 285, 288
– und mentale Modelle 118ff
– Ähnlichkeit von kausalen Rollen 119
– und Qualia 214, 239, 260

Ganzheiten
– phänomenale 36, 95
Gedächtnis 69
Gedankenketten 64
Gegebenheit
– der Gerichtetheit 31
– phänomenale Qualität 65, 67, 73, 77ff, 246ff, 249
– Selbstgegebenheit 268
Gegenwart
– aktuelle, äußerer Eigenschaften 75
– „Gegenwartsfenster“ 77, 102
Gehalt
– mentaler Phänomene 17ff, 268
– propositionaler 17f
– intentionaler 17f, 25, 57, 67, 134, 198, 206, 215f, 246
– kausaler 19ff
– phänomenaler 20, 25, 64f, 67, 73, 126ff, 206, 212ff, 238ff, 246, 258ff, 273ff
– repräsentationaler 25ff, 50, 215
– qualitativer 31f, 105, 113ff, 142ff, 167, 212ff, 258ff, 273ff
– mentaler Repräsentate 52
– Funktionalität von 52, 67, 74
– und Systemkontext 53
– Verschmelzung verschiedener Formen 59, 86
– Konstanz durch Musterwiederholung 64
– Nomizität 69, 118
– Kovarianztheorie 69
– von Simulaten 70ff
– introspektive Zuschreibung 83, 255f
– höherer Ordnung 94
– von mentalen Modellen 121ff
– epistemischer 132, 170, 185
– Plastizität/Variabilität 138, 163
– Diskontinuitäten 140ff
– Dissoziation von phänomenalem und intentionalem 141ff, 206, 258ff
– phänomenale Lokalisierung 163
– nur sequentieller 164f
– subjektiver/objektiver 207
– Theoriefähigkeit 239
– semantischer 251
– „Verobjektivierung“ von subjektivem 254ff
– Determination von unten 285
Geistbegriff
– Revisionen des 15
– metaphysische Phase 20, 50
General Problem Solver 243, 287
Gesamtzustände
– repräsentationale 62, 133, 135ff, 174, 225f, 254, 283
– abweichende 177ff
– simulationale 194
– luzide/transparente 199f
– Plastizität 287ff
Geschichtlichkeit 23
Gesetze
– strikte 19
– psychophysische 26
Gestaltkriterien 106
Gewißheitserlebnis
– allgemein 74, 211
– Körper-Selbst 157ff
Gipfelerlebnisse 136
Gleichzeitigkeitshorizont 77

Halluzinationen 56, 65f, 113, 116, 137ff, 156, 166, 177ff, 185ff, 259
Handeln
– inneres 69, 119, 137, 152f, 155, 172, 179, 243, 288
– externes 121, 137, 152f, 155, 243, 288
– im luziden Traum 195ff
– rationales 286, 289f
Hemi-Neglekt 139ff
Heterophänomenologie 128
Higher-Order-Thoughts 55, 65, 98
Hintergrundbedingungen
– physische, psychischer Zustände 16
Hintergrundwissen
– implizites 28
Homogenität
– repräsentationaler Gesamtzustände 61

– zeitliche, von Bewußtsein 91, 102
Homomorphie
– partielle relationale 62, 110, 280
Homunkulus
– hinter den Augen 9
– und innere Aufmerksamkeit 55, 262
– und innere Bilder 58, 99
– und semantische Transparenz 63, 257
– der Psychoanalyse 201
– im System 243
– transzendentaler 255
Humanchauvinismus 167
Hybridsysteme 111
Hypnose 54, 182ff, 191

Ich
– reines 127, 228, 256
– phänomenales 154f, 257, 266ff, 280
– Biologizität 162f
– Auflösung 182ff, 217, 281
– Ich-Welt-Grenze 184
– Grenze der Welt 228f, 288
– transzendentales 230, 255
– „Subjektgebrauch“ 252ff
– Reduzierbarkeit 271
– Opazität 286
Idealismus
– transzendentaler 228
Ideen 50
Identität
– psychische 21ff, 191ff, 216ff, 227, 264ff, 285
– Aussagen 40, 222ff, 230ff, 251ff, 271f, 280
– Selbstmodellerzeuger/Sprecher 252ff
– monologisierende 255f
– erlebte 267
Identitätskriterien
– für mentale Zustände 17
Identitätstheorie 24, 80, 174, 214, 237, 269, 271f
Imagination
– empathische 45, 214f
– als Erzeugung unterdeterminierter Datensätze 67
– und Analytizität 167, 226
– intendierte 225f, 255, 281ff
Indikatoren
– indexikalische Ausdrücke 39, 81f, 230ff
– digitale 82ff
– analoge 82ff
– indexikalische Gedanken 83, 230ff
– Propositionen für mentale Modelle 133
– und aktives Selbstmodell 152ff
Individuen
– phänomenale 55, 107ff, 117f, 236, 269
– nicht-physische 232
Inexistenz
– mentale 127
Infallibilität 250
infinite Regresse 63
Information
– Begriff der 49, 127ff
– phänomenale 76, 81f, 219ff, 259ff
– visuelle 87
– Informationsdichte 136f
– als Schleier 286
informationale Einkapselung 79
Informationsverarbeitung
– biologische 48, 80, 95, 287
– und psychische Episoden 49
– und psychologische Eigenschaften 49, 245f
– subsymbolische 50, 175, 180
– neuronale 57, 77, 86, 107, 117, 138ff, 227, 268
– nicht-lokalisierte 61
– zeitliche Quantelung 77, 90, 102
– und Intentionalität 127ff
– als Hintergrundannahme 206
Informationsverarbeitungsansatz
– und Objekthaftigkeit von Bewußtseinsinhalten 57
Inkommensurabilität
– nomologische, des Mentalen 25, 32, 48
Inkorrigibilitätsthese 269
Innenleben 16, 87, 212, 289
Innenperspektive 213
Innerlichkeit
– phänomenale 20, 37, 85, 96, 211, 261, 265ff, 269f, 285, 291
– physikalische/repräsentationale 269f, 290f
Instantanität
– allgemein 67, 73, 76ff, 100ff
– des Körpergefühls 158
Instantiierungserklärungen 96, 100, 120, 208
Intelligenz
– künstliche 28
Intensitätsdarstellung 61, 82
intentional stance 128, 215f
Intentionalität
– des Mentalen 17ff, 54
– durch Informationsverarbeitung 47ff, 109
– „Pfeil der Intentionalität“ 107, 128
– und Mechanismus 125f
– durch mentale Modelle 127ff
– „invertierte“ 128
– „interne“ 129
– „klassische“ 136
– „mereologische“ 151ff
– als inneres und äußeres Handeln 152f

– instrumentalistischer Begriff 215f
Interdisziplinarität
– der analytischen Philosophie des Geistes 9, 20
– der Diskussion 13
Interessen 53, 134, 152ff, 158, 162, 184, 206f, 244, 263f, 270, 274f
Intermediate-Level Theory 94
Intersubjektivität
– moralische 26
– theoretische 277
Introspektion
– allgemein 35, 38, 55, 74, 77, 89ff, 180, 213ff, 225, 239, 286
– prozessorientierte 56
– Berichte 129, 255f, 261ff
– als Meta- und Selbstmodellierung 159, 166
– KI-Systeme 276
Introspizierbarkeit 55, 96ff
Intuitionen
– allgemein 31, 87, 101, 137, 142, 263, 274
– essentialistische 36, 81, 233, 271f
– cartesianische 39, 76, 109, 180, 213ff, 227ff, 256ff
– historische Plastizität 85
– theoretische 87, 108, 127, 163
– modale 214, 237
– epistemische 219
– „Kontingenz-Intuition“ 223, 246ff
– „Zentriertheits-Intuition“ 237
– „Unteilbarkeits-Inuition“ 246ff
Isomorphismen
– allgemein 154, 159
– funktionale zweiter Ordnung 111
– höherstufige 120, 268
– kausal-teleologische Einbettung 121f, 132ff

Jetzt
– phänomenales 77
– subjektives 102, 180

Kategorienfehler 20
kausale Geschlossenheit
– der physikalischen Welt 19
kausale Rolle
– des intentionalen Gehalts 18
kausale Verhaltenserklärungen
– und intentionaler Gehalt 18
– und Rationalität 18, 290
Kausalität
– transphysikalische 25
KI-These, starke 275
Klarträume 67
– siehe Träume, luzide
Knowledge by Acquaintance 221
Knowledge by Description 221
Knowledge-Argument 76, 218ff, 239
Kognitionstheorien
– empirische, und Naturalisierung der Erkenntnistheorie 18
kognitive Nische 189
kognitive Penetrabilität 58, 108
Kohärenz
– von Ableitungen 125f
Komplexität 89, 206, 286f
Konfabulationen
– externe 143f
– interne 182, 186, 262
Konkretheit
– unseres Innenlebens 22
– phänomenale Qualität 64, 264, 288
– von Qualia 84
– und Biologizität 264
Kontext
– und repräsentationaler Gehalt 53
– psychischer 82
– Kontextsensitivität von Begriffsstrukturen 125f
– kausal-teleologischer 133ff
– sozialer 134
Kontinuitätsillusion 64
Kontrafaktizität
– Situationen 66, 70ff
– Ereignisfolgen 69
– Kompetenz für Kontrafaktizität 112
Kontraintuitivität 30
Körper
– Körpergefühl 157ff, 177ff, 267f
– Körperschema 157ff, 168, 177ff, 187, 267f, 288
– phänomenales Körper-Selbst 157, 177ff, 264, 278, 288
– Körpermatrix 267, 288
Korrelationen
– psychophysische 19
– gesetzesartige 25, 96
Kurzzeitgedächtnis 95, 276

Language of Thought 125f, 133, 268
Lebenswelt
– lebensweltliches Selbstverständnis 42, 237, 260ff
– mesokosmische 96
– naiver Realismus 109, 242, 249
Leib-Seele-Problem
– als Problemfeld 17ff
– moderne Diskussion des 19, 216
– Varianten 49, 93f
– allgemein 54, 69, 95, 118, 210
– vorphilosophisches 161
Leiberleben 156ff, 185f, 278
Leidensfähigkeit 27

Liberalismus-Chauvinismus-Dilemma 93
lingua mentis 127
Lokalisierung
– phänomenale 114
lucid dream
– siehe Träume, luzide

Makro-Eigenschaften 95, 169
Makro-Simulate 135, 185
Manien 136, 191
Maschinen
– virtuelle 165, 199, 287f
Maschinensprachen 80
Materialismus
– nicht-reduktiver 32, 89, 172
– eliminativer 45, 172, 212, 214
– emergentistischer 89
– minimaler 96, 202
– „cartesianischer" 128
Meditation 56, 101, 237
Meinigkeit
– phänomenale Qualität der 23, 109, 168, 216, 236, 253, 265ff, 280, 285
– unbewußte 267
– Epistemizität 268f
mental
– Zuschreibungskriterium 54
– hinreichende Bedingungen 56
Mentale Modelle
– allgemein 50f, 76, 104ff
– Metamodellierung 84, 114ff, 270
– als naturalisierte intentionale Objekte 107, 128ff
– relationale Struktur 110ff
– Analogizität 112
– Einbettbarkeit 114, 123ff, 255, 266
– Variablenfreiheit 115ff, 244
– funktional aktiv 118
– Gehalt und kausale Rolle 119
– Diskurs-Modelle 123
– empirische Plausibilität 124
– als Instrumente 128ff, 166ff, 235f
– aktive 135
– ohne räumliche Parameter 161
– von Propositionen 163
– „Bildhaftigkeit" 169f
– Typologie 169
– Eigenschaften 251, 266
– privilegierter Zugang 270
– ontologische Neutralität 276
Mentalesisch 125
Mentalität
– als dispositionale Eigenschaft von Datenstrukturen 54, 272
– Zuschreibungskriterien 85, 272
– analyseresistente Elemente 88f
– von internen Selbstrepräsentaten 152, 159
Metakognition 55
Metarepräsentation
– als logisch unendlicher Prozess 57
– und Simulation 63,65
– und Präsentation 77ff, 83
– subsymbolische und Bewußtsein 87ff, 98
– und mentale Modelle 114ff,136ff, 152, 270
– Störungen 137ff
– des Selbstmodells 159, 165f, 252ff
– und Gegebenheit 248f
Mikro-Ereignisse 96
Mikro-Makro-Intuition 96
Mind-Mind-Problem 65
Mittelpunkte
– räumliche/phänomenale 222ff
Modal argument 214
Modellgeneratoren 111, 244
Module 108
Monologe
– innere 63, 65
motorische Funktionen 159f
Multimodalität 59, 102, 106ff, 114ff, 242f
Multiple Personality Disorders 190, 217, 221, 237, 266, 278
Multirealisierbarkeit
– funktionaler Zustände 52
– mentaler Modelle 121

Nachbilder 76
Nachdenken 76
Naturalismus
– als Erklärungsstrategie 9
– und die Innerlichkeit subjektiven Bewußtseins 20
natürliche Arten 27
– innere 35
natürliche Zeichen 118
Neglekte 56, 137ff, 177ff
Neuro- und Kognitionswissenschaften 13, 21
neuronale Netze
– allgemein 48, 52, 117, 125f, 247, 275
– selbstorganisierende 91
NMDA-Antagonisten 156
NMDA-Rezeptor-Komplex 91
Noema 127
Noesis 127
Nominalismus 239
non-veridical seeing 142

objective self 58, 152, 155, 183, 211, 221ff, 230, 256ff, 281ff (siehe auch Selbst, objektives)
Objektbildung
– in der Logik 23, 57

– im Gehirn 57, 98, 229, 288
– im Cartesianismus 88f
– in externen Codes 117, 229, 233, 256f
– als Subjektbildung 168, 284ff
Objekte
– moralische 26, 153
– transmundane 77
– intentionale 107, 128
– mentale 128, 233, 257
– innerweltliche 210f
– private 239, 269
– repräsentationale 283
Objekthaftigkeit
– von Bewußtseinsgegenständen 57, 64, 73, 229, 249
– von Elementen des Weltmodells 109, 249, 264
Objektivität
– Kompetenz für 45, 287f
Objektivitätsideal 20
Ökologie
– „epistemische“ 16
Ontologie
– psychische 33
– phänomenale 35
– psychologische 271f
ontologische Neutralität 213, 276
Opazität
– selbstreferentielle 63
– von Subjektivität 173
Organe
– abstrakte 59, 152, 159ff, 173, 189, 245, 285, 290 (siehe auch *Abstrakta*)
Out-of-body-experiences 187ff
Output-Entkopplung 68

Parallel-Algorithmen 93, 288
Performanzebene 29
Person
– Begriff 132, 166, 172, 216
– phänomenale Person 132, 166, 171ff, 264
– Identität der phänomenalen 180
– Traumfiguren 198, 278
– als Erkenntnisinstrument 256ff
– fiktive 266
Persönlichkeit
– multiple 190ff
– affektives Profil 190f
– Spaltung 237
Perspektive
– der ersten Person 23, 128, 216
– perspektivische Fakten 23, 221ff, 281ff
– quasi-objektive 46
– Perspektivenwechsel 185, 229ff, 254ff, 281ff
– „innere Außenperspektive“ 254, 256f
Perspektivität
– in der Phänomenologie des visuellen Sinnes 9
– des phänomenalen Raums 9, 168ff, 248, 250, 277, 281ff, 288f
– des Objektbildungsvorgangs 23
– Struktur von Repräsentationsräumen 96, 168ff, 172, 244, 281ff
– multiple/alternierende 189ff
– phänomenalen Bewußtseins 205ff, 211, 232, 264ff
– als Erlebnis 282
– Abschwächung 287f
phänomenale Familien 83
phänomenale Realität
– Struktur der 60ff
– Traum als Innenraum 187, 200
Phänomenalität
– als Eigenschaft psychischer Zustände 34
Phänomenologie
– objektive 44, 289
Phantomglieder 160, 187
Phenomenal Mind 65
Phenospace 243
Philosophie des Geistes
– analytische 15, 20
– als Metatheorie der Einzelwissenschaften 14f
– als begrifflicher Kommentar 15f
philosophische Oneirologie 200
Phosphene 113
Physikalismus 210f, 214, 220
Planung 65ff, 69
Pneuma 187
prana 187
Präsentandum
– des Selbstpräsentats 157
Präsentate
– nicht simulationsfähige 66f
– output-entkoppelte 68
– phänomenale Ahistorizität 77
– reine 114
– nicht multirealisierbar 121
– funktionales Profil 260ff, 275
– Plastizität 260, 287f
– propriozeptive 267
Präsentation
– interne 62
– „purer Präsentationsaspekt“ 68, 156ff
– mentale 72ff, 259ff
– Fehlpräsentation 75f
– phänomenale Präsenz 264
– ohne Metarepräsentation 259
Primitivum
– das Subjekt als logisches 38

Privatheit
– des Mentalen 20, 210, 218, 269f
Produktivität 125f
Profil
– behaviorales 28, 219
propositionale Formate 70, 125ff
Propositionen
– mentale Modelle von 133, 265, 276
Prosopagnosien 138
Prototyp
– theoretischer des phänomenalen Repräsentats 105ff
– und Begriffsbildung 131ff, 180
Prozesse
– epistemische 18
Prozessualität
– des Mentalen 41, 50, 233f
– von Gedankenketten 64f
– kausaler Antezedentien 78
– von mentaler Modellierung 117f
– psychophysische 153
– des *Blick von Nirgendwo* 229
– erlebte 275
Pseudo-Halluzinationen 67
psychoaktive Substanzen 67, 86, 116, 182ff, 221, 231, 237
Psychologie
– universelle 16, 43, 92
– empirische 21, 37, 41, 260ff
– natürliche Psychologen 35, 257
– naturalistische Psychologen 35
– personale 194
– „Selbstmodell“ als Erklärungsbegriff 208
– philosophische 235
Psychosemantik 134, 231
Psychosen 67, 181
Psychotherapie 55

Qualia
– allgemein 32f, 36, 211ff, 238ff, 278ff
– als phänomenale Atome 36, 80f
– Irreduzibilität 36, 79ff
– und mentale Präsentation 72ff
– ahistorischer Charakter 74
– dualistisch-metaphysische Interpretation 74
– als Informationsträger 75, 259ff
– transzendentale 77
– Elimination 79
– als Epiphänomene 79
– vortheoretischer und technischer Begriff 81, 260ff, 270
– Selbstzuschreibung 83, 260ff
– als repräsentationales Phänomen 84, 270
– „intellektuelle“ 163
– Plastizität 260
– funktionales Profil 260ff, 275
– nicht-subjektzentrierte 263
– emotionale 274f
Qualitäten
– subjektive 20
– taktile/haptische 73
– Idiosynkrasien 86
– phänomenale 137

rapid eye movements 196f
Rationalität 133, 214, 289f
Raum
– phänomenaler 248
– epistemische Geschlossenheit des inneren 269f
Rauschen
– höherstufiges 114
Realisierung
– zu liberaler Begriff 53
– physische 80, 121
– neurobiologische 97
Realismus
– naiver 108, 249, 255f, 283f
– intuitiver 109, 255f
– wissenschaftlicher 215, 277f
Realitätsmodelle
– allgemein 62, 114
– multiple 136
– dezentrierte/nicht-zentrierte 184ff, 257, 264ff, 281ff
– luzide/transparente 199f
– generelle/benutzerfixierte 244f, 261ff, 277f, 281
– primäre/sekundäre 255f
– afunktionale 265
– soziale/mentale 289
Reduktion
– physikalistische 28
Reflexbogen 68
réflexion virtuelle 99
Reifikationen
– sprachliche 117, 256, 282
– metaphysische 229, 233
– phänomenologische 234
– psychische 236
Relationalität
– mentaler Zustände 22f, 218
– Struktur der Welt 137
– von Subjektivität 221ff
Relationen
– computationale 33
– Subjektivität als epistemische 40, 173ff, 220f
– epistemische 130ff, 173ff
– extraorganismische 245
religiöse Erfahrungen 136

REM-Phasen 196f
Repertoire
– behaviorales 69
– *inneres behaviorales* 69
Repräsentanda
– allgemein 34, 51ff
– aktuell gegebene 113
Repräsentate
– allgemein 34, 51ff
– im Gegensatz zu Repräsentationen 51
– subsymbolische 52
– als innere Werkzeuge 53, 65ff
– als extern erlebte 56
– imaginale 58, 67
– phänomenale, idealisierter „theoretischer Prototyp“ 60ff, 105ff
– dynamische 61, 64
– multimodale 61, 64
– prototypische 61, 64
– relationale Struktur 62ff,
– als Konstrukte 63, 65, 72, 283
– input-unabhängige 65ff, 68
– output-entkoppelte 68f
– höherstufige 69, 101
– propositionale/piktoriale 80
– propositionale 83, 125ff, 131, 164ff, 235
– Bewußtsein als Makro-Repräsentat 103
– distribuierte 119
– funktional aktive 134
– nicht-sensorische zweiter Ordnung 161
– benutzerfixierte 243ff, 261ff, 277f, 281
Repräsentation
– theoretische 37, 96
– mentale 37, 47ff,
– naturalistische Theorie der mentalen 47ff
– von Eigenschaften der Welt 48
– interne 48ff
– Geschichte des Begriffs der 49f
– logische Struktur der Beziehung 51ff
– Begriff der mentalen 51ff
– externe 54
– nicht-mentale interne 54
– mentale als psychisches Phänomen 57
– Bildhaftigkeit 58
– imaginale 58
– Biologizität 59ff
– als abstrakte Organe 59ff, 152ff
– logisch mögliche 59ff
– phänomenologisch realistische 59ff
– als Sonderfall von Simulation 70ff
– Fehlrepräsentation 154
– analoge 154ff
Repräsentationalität
– als Eigenschaft psychischer Zustände 34
– des Erkenntnisbegriffs 50
– und Funktionalität 69
Repräsentationsräume
– zentrierte 245ff
– perspektivische 256f
Repräsentationssysteme
– natürliche 58, 103, 110, 260ff, 289
– künstliche 79
– und Meta-Simulation 101
– intersubjektive 84, 262, 277f, 289
– *scientific communities* 96, 133, 199, 260ff, 277f
representatio 49f
res cogitans 154f, 247, 276
res extensa 160, 179
Risikoabschätzung 71
ruach 187
ruh 187

Schemas 123f
Schizophrenie 135, 181, 185, 191, 272
Scripts 123f
second-order-volitions 275
Seelenblindheit 138f
Seinsgewißheit 136
Selbst
– phänomenales 9, 159ff, 266ff
– allgemein 22, 210f, 236
– als fiktiver Fixpunkt 23
– naturalistische Theorie des 26
– objektives 45, 227ff, 255ff, 282
– phänomenales 155ff, 283f
– denkendes 161
– multiple 177ff
– halluziniertes 177ff, 185
– als Originator von Gedanken 180
– „ozeanische Entgrenzung“ 182ff
– fiktives 185
– Selbstverlorenheit 186f
– Spaltung 192ff
– virtuelles 242f
Selbst-Symbol 279
Selbstähnlichkeit 153ff, 173, 290
Selbstbewußtsein
– als Ichbewußtsein 99
– aktueller Inhalt 171
– Logik/Psychologie 175
– als psychologische Eigenschaft 204ff
– qualitatives 272ff, 290
Selbsterkenntnis 154, 170, 246, 258, 279, 286, 291
Selbstmißverständnis
– naiv-realistisches 174, 186, 243f, 257ff, 283f, 286, 290
Selbstmodell
– Begriff 158ff, 204ff, 243ff
– Genese 159ff, 189
– geometrisches 160f

– repräsentationales Fundament 160ff, 288
– nicht-räumliche Elemente 161ff
– metakognitives 161
– emotionales 162f, 190, 193, 262
– „nicht-weltliche“ Elemente 164f
– propositional darstellbare Elemente 165
– kognitiv-intellektuelles 168, 181f
– als komplexes „Bild“ 169
– Funktionalität 170, 189ff, 207, 282f
– Gehalt als Referenz 175f
– deviante 177ff
– Multimodalität 179
– gespeicherte 180
– multiple 190ff
– Benutzerfixierung 243ff, 261ff, 277f, 281
– öffentliche Eigenschaften 251ff
– Distanzierung 257ff, 288
– Variablenfreiheit 277
– Plastizität 283
– als Organ 285
– als Menge funktionaler Zustände 285, 288
– als autonomes Subsystem 288
– kulturelle Einbettung 289
Selbstmodell-Theorie der Subjektivität 204ff
Selbstmodellerzeuger
– künstliche 170
– Menschen 244, 286
Selbstmodellierung
– interne 10
– mentale 47, 250
– „kontrafaktische“ 156
– als mentaler Agent 166
– Repräsentation/Simulation 166
– als soziales Wesen 171, 189ff, 289
– und Selbstreferenz 175f
– abweichende Formen 177ff
– als Träumer 202f
– als Internalisierung sozialer Aktivitäten 208
– als fremde Person 255f
– als Subjektivitätskriterium 272f, 279ff
– endloser Progress 283
Selbstpräsentation 156ff
selbstreferentielle Opazität 154f, 173
Selbstreferenz
– monologische 40, 235, 249f, 254f
– und Selbstsimulation 169ff, 249f, 279f
– und perspektivische Fakten 225ff, 251ff
– als innerpsychische Orientierung 234
Selbstrepräsentate
– allgemein 151ff
– als Konstrukte 155, 283f
– digitale/propositionale 164ff, 175f
Selbstrepräsentation
– mentale 128, 151ff
– mereologische 153
Selbstsimulation
– allgemein 155ff
– und Selbstreferenz 169ff
– pathologische 182
– afunktionale 186, 248
– fiktive 248
Selbsttransparenz
– des Bewußtseins 31, 173, 199
– des Subjekts 269f
– repräsentationale 283f, 287f
Selbsttranszendenz 281
Selbstwissen
– analoges/internes 165, 169ff
Selbstzuschreibung psychologischer Eigenschaften
– introspektive 213f, 255f
– monologisierende 175, 254f
– unter der Hinsicht der Internalität 176, 251ff, 255
– unter der Bedingung der Identität der die Repräsentate produzierenden Systeme 176, 250ff
self-locating thought 45
semantische Transparenz 63f, 115ff, 155, 173f, 257
Sententialismus 125, 133
Signalaspekt 67f (siehe auch *Signale*)
Signale
– analoge 61, 169
– Stärke 61, 82, 156
– interne Quellen 67, 75, 157ff, 203, 259
– Signalcharakter von Qualia 75ff, 259
– „Signalaspekt“ 113, 139ff, 157f, 264, 264, 288
Simulate
– kausale Geschichte 68
– perzeptuelle Komponente 72
– Variablenfreiheit 117
– „motorische“ 119
– propositionale 126
– Prozessualität 130
– afunktionale 181f, 245
Simulation
– mentale 47, 62f, 65ff, 130ff, 247
– genetische und logische Perspektive 67ff, 70
– Sonderfälle 71
– intendierte 72
– phänomenale 75ff, 113
– mentale Meta-Simulation 101
– bewußte 101
– und mentale Modelle 112ff
– assoziierte mentale 112
– emotionale 158, 162f

– von nicht-zentrierten Realitätsmodellen 255ff
– von Fremdsystemen 278
Skotom 141, 259
Somnambulismus 54, 69
Speziezismus
– psychologischer 16
– ethischer 27
spiritus 187
split-brain-Patienten 31, 217
Spontanität 100
Sprachzentren 95
Sprechen 70, 275
Sprechergemeinschaft 276
Standard-Kausalketten 70
Standpunkt
– phänomenaler 25, 265
– subjektiver 222ff
Sterblichkeit 285, 290
Subjekt
– psychischer Zustände 22ff
– erkenntnistheoretische Autorität des 37, 251ff, 260ff
– psychologisches, als Initiator 72, 286
– psychologisches 77, 81, 168, 198ff, 210f, 232f, 245-284
– Einheit des 88f
– Rückwendung in sich selbst 99
– Identität und wissende Selbstbeziehung 99
– Identität von Subjekt und Objekt 100
– moralisches 153, 289
– soziales 153, 289
– Selbstgewißheit des 154f, 223
– innerer Erlebnisse 159, 244
– logische, unterschiedliche, 173f, 205, 48
– luzides 203
– Relationalität 221ff
– Identität 221
– innerpsychische Orientierung 234, 254ff
– Einzigartigkeit 245, 278
– intellektuelles 249, 258, 275
– transzendentales 258
– Immanenz 290
Subjekt-Argumentstelle 23f
Subjektbegriff
– phänomenologische Säulen des 21ff
– psychologisch-philosophischer 22, 92, 250ff
– ethische Aspekte 26f
– metaphysischer 74, 78, 128, 164, 240, 269f, 286ff
– essentialistischer 189, 248, 271ff
– nicht-psychologische Elemente 208, 250ff, 270ff
– anthropozentrischer 263f, 274
– logische Primitivität 270ff
– schwacher/starker 272ff
– Naturalisierung 285
Subjektgebundenheit mentaler Zustände 23, 26f
subjektive Bewußtseinszustände
– und ihr phänomenologischer Reichtum 16
Subjektivität
– als Problemfeld 17f, 38ff, 96
– psychologische 26, 41ff, 174ff
– als psychisches Phänomen 26, 210ff, 225f
– Reduktion auf Repräsentation 27
– als epistemisches Phänomen 37, 99, 268ff
– logisch-semantische 38ff
– erkenntnistheoretische 40ff
– Theorie der 54
– als relationale Eigenschaft 62, 69, 221ff
– nicht-relationaler Aspekt 88ff
– Reflexionsmodell 97ff
– und Selbstmodellierung 151ff, 272
– leibgebundene 161
– nicht-diskursive 171ff
– Opazität 173
– sozial konstituierte 175f, 208, 274ff
– Zerfallsbedingungen 183, 272
– von repräsentationalen Gesamtzuständen 205f
– von mentalen Modellen 205f
– künstliche 207, 272ff, 278
– logische 208
– grammatische 208
– als Faktum 222ff
– partikulare 222ff
– „propositionale“ 275
Subjektzentriertheit mentaler Zustände 23, 57, 167, 216, 236ff, 267f, 272
Substanz-Attribut-Denken 118
Substanz-Attribut-Relation 211, 217
Substanzentheorie
– cartesianische 18f
Supervenienz 25, 95f, 172, 202, 256
symbolische Atome 52
Symbolmanipulation 70, 125ff
Synästhesien 86f, 183
syntactic engines 111, 125ff
System
– intentionales 28, 215f
– phänomenales 28, 34
– als Ganzes 52, 151ff, 204ff, 243, 250, 253
– postbiotisches 92, 275
– Möglichkeitsraum phänomenaler Systeme 103
– künstliche/natürliche 134, 212, 261, 271ff, 276
– Systemperspektive 156
– physische Grenze 160, 279

– träumendes 203
– intendierte Klasse 207, 244, 271ff, 276
– funktionale Architektur 246, 251
Systematizität 125f
Szientismus 14, 84

Tagtraum 56, 63, 65, 155, 282
Tatsachen
– aus der Perspektive der ersten Person 37, 282ff (siehe auch *Erste-Person-Fakten*)
Taxonomie
– wissenschafliche 29, 246, 253
Teleofunktionalismus 53, 93
teleologische Zusatzkriterien
– für funktionale Zustände 53, 66, 121f, 264
– für Bewußtsein 72
– für Qualia 77ff
– für Selbstmodellierung 152ff, 276
Theater Model of Mind 50
Theorie des Mentalen 21
Theoriebildung
– als kognitive Leistung 69
Theorien des Geistes
– historisch 13f
– mythologische 13f
– theologische 13f
– philosophische 13f
– postmetaphysische 15, 168
– naturalistische 18, 29, 31, 35, 97f, 100, 120, 152, 154, 284ff, 290f
– anti-naturalistische 34, 74, 84, 87f, 238
– materialistische 57
– idealistische 57, 100
– „informationalistische“ 57
– metaphysische 73, 86, 194
– esoterische 86
– egologische/substantialistische 88f, 194
– generelle 91, 167
– *Representational Theory of Mind* 94, 284ff
– klassische 138, 204
– mentale Realitätsmodelle als „nicht-öffentliche“ 143
– „mythische Prototheorie“ 187f
– platonisch-cartesianische 243f
top-down-Analyse 80, 120
Totales *Locked-In*-Syndrom 98
Transduktoren 108
Träume
– normale 62, 65ff, 137ff, 155f, 166, 186f, 259
– luzide 177, 188, 194ff, 278
– Funktion 201
Turing-Maschinentafeln 28
Typologie
– mentaler Repräsentate 59
– abweichender phänomenaler Zustände 182ff

Umwelt
– als Netzwerk kausaler Relationen 62, 245, 275
– aktuelle 66
– biologische 134, 152, 158
– soziale 171, 275
– Interaktion 288
unbewußte zerebrale Initiativen 72, 286
Unhintergehbarkeit
– phänomenale 141
Universalien
– mentale 43, 91
Unteilbarkeits-Intuition 30, 36, 246ff
Ursachenketten 18

Variablen 115, 277
Verallgemeinerungen
– gesetzesartige 19
Verhalten
– rationales 29, 289f
– diskriminatorisches 87
– Erklärungen 246
– Repertoire 285
Verkörperung 156ff, 264
Verursachung
– von oben 50, 55, 202
Vigilanz 139
virtuelle Realitäten 242f
virtuelle Selbste 242f
Visionen 113
von-Neumann-Kognitivismus 126
von-Neumann-Maschinen 52
Vorstellung 49f
Vorstellungsbilder 58, 67

Wahrheit
– und innere Zustände 18, 167
– „psychologistischer Wahrheitsbegriff“ 123ff
– aus der ersten Person Singular 221ff, 231
– doppelte Wahrheitsbedingungen 231, 251ff
– nicht-triviale 253ff
Wahrheitswerte 129
Wahrnehmungen 33, 122
Wahrnehmungserlebnisse 33, 219
Wechselwirkungstheorie 19, 89
Welten
– mögliche phänomenale 71, 271f
– biologisch wahrscheinliche 71
– abweichende phänomenale 87, 281
– phänomenale 135ff
– subjektiv erlebte 138ff

– logisch mögliche 248, 271f
Weltmodelle
– allgemein 98, 101, 103, 114, 124ff, 130ff, 135ff, 154, 160
– phänomenal deprivierte 139ff
– zentrierte 166ff
Widerständigkeit
– phänomenale 73
Wie es ist, ein X zu sein 166f, 201, 211ff, 278ff
Wirklichkeit
– innere 68
– Wirklichkeitsmodelle 124ff, 135ff, 283f
Wissen
– über die Welt durch mentale Zustände 18, 135
– inneres 82, 210ff, 232, 239ff
– Faktenwissen 82
– deklaratives Meta-Wissen 98
– inneres propositionales 98
– Wissensstrukturen 123f
– digitales 125ff, 279
– analoges 125ff, 154ff, 268ff, 279
– kategoriales 126
– durch Repräsentation 127ff
– durch Sätze 130ff, 220f
– intersubjektives 133, 171, 277
– nicht-diskursives 133f, 171ff, 220f, 261, 270
– subjektives 133, 218ff
– funktionales 134
– außerhalb von Sätzen 172
– nicht-inferentielles 217
– qualitatives 219ff
– internes *als* externes 228
– privates 221, 269f
– meta-psychologisches 262, 282
– implizites Hintergrundwissen 275
– unvollständiges wissenschaftliches 252ff, 277f
– biologisches 282

Zeitbewußtsein 102, 153
Zeitlichkeit
– von Prozeß und Gehalt 77
– temporale Codierung 106
Zentriertheit
– von Realitätsmodellen 138, 205ff, 222ff, 237, 248, 265, 271, 289
Zentrum
– des visuellen Erlebens 9
– phänomenales 168, 237
Zielhierarchie 163, 274
Zielvariablen 95, 162, 274
Zielzustände 62, 65, 155, 162
Zustände
– qualitative 20
– wissenschaftliche Taxonomie psychischer 29, 246, 253, 275
– subjektive 33
– von Maschinen 33
– mentale, nicht-repräsentationale Aktivierung 70
– psychische 85
– logisch unmögliche phänomenale 85
– mentale, relationale Eigenschaften 94
– funktionale 120, 245, 285, 288
– hypnagoge 182
– personale 206, 285
– von Systemen 206, 268
– intentionale 215f
– phänomenal fremdsubjektive 217
– als Objekte 233
– des Selbst 236
– präsentationale 260
Zustandsklarheit 67, 155, 199ff
Zustandsraum
– neuronales Netz 247
– phänomenaler 287ff
Zustandsspezifität der Philosophie 198